

Tartu Ülikool
Loodus- ja täppisteaduste valdkond
Matemaatika ja statistika instituut

Hanna-Liisa Soll

Haiglapatsientide arvu prognoosimine ARIMA tüüpi mudelitega

Matemaatika ja statistika õppekava
Matemaatilise statistika eriala
Magistritöö (30 EAP)

Juhendajad: Krista Fischer, PhD
Raul Kangro, PhD
Merje Tikk, MD (TÜ Kliinikum)

Tartu 2019

Haiglapatsientide arvu prognoosimine ARIMA tüüpi mudelitega

Magistritöö

Hanna-Liisa Soll

Lühikokkuvõte. Käesoleva magistritöö eesmärk oli prognoosida igakuist Tartu Ülikooli Kliinikumi haiglapatsientide arvu neljal ravierialal kuni üks aasta ette. Vaadeldi otorinolarüngoloogia, pediaatria, sisehaiguste ning üld- ja plastikakirurgia erialasid. Põhieesmärgiks oli leida sobivad ühemõõtmelised ARIMA tüüpi prognoosimudelid. Lisaks prooviti prognoose täpsustada, kasutades lineaarset regressiooni ARIMA tüüpi vigadega, võttes regressoriks Eesti Haigekassa poolt aastaks tellitavate ravijuhtude arvud vastavatel erialadel. Iga eriala jaoks valiti parim ühe- ja mitmemõõtmeline mudel ning võrreldi nende täpsust aastaste prognooside ruutkeskmiste vigade põhjal. Saadud mudelite abil leiti prognoosid valimiväliseks aastaks 2018 koos prognoosiintervallidega.

CERCS teaduseriala: P160 Statistika, operatsioonianalüüs, programmeerimine, finants- ja kindlustusmatemaatika

Märksõnad: aegridade analüüs, prognoosimudelid, haiglapatsiendid

Predicting Inpatient Counts Using ARIMA Type Models

Master's thesis

Hanna-Liisa Soll

Abstract. The aim of this master's thesis was to predict monthly inpatient counts in four specialities of Tartu University Hospital for up to one year in advance. The specialities considered were otorhinolaryngology, paediatrics, internal medicine, and the combination of general and plastic surgery. The emphasis was on finding appropriate univariate ARIMA type prediction models. In an attempt to improve prediction accuracy, linear regression with ARIMA type errors was implemented using the yearly number of treatment cases in the corresponding speciality commissioned by Estonian Health Insurance Fund as the regressor. For each speciality, the best univariate and the best multivariate model were selected and compared based on yearly root mean square errors. The monthly predictions of these models for the out-of-sample year 2018 were given along with prediction intervals.

CERCS research specialisation: P160 Statistics, operations research, programming, actuarial mathematics

Keywords: analysis of time series, predictive models, hospitalized patients

Sisukord

Sissejuhatus	5
1 Metoodika	6
1.1 Aegridade teooria mõisteid ja omadusi	6
1.2 Ühemõõtmelised mudelid	9
1.3 Mitmemõõtmelised mudelid	12
1.4 Mudelite sobivuse ja täpsuse analüüs	13
2 Andmete kirjeldus	17
2.1 Haigete arvu kujunemine	17
2.2 Haigete arvu andmestike moodustamine ja korrastamine	18
2.3 Otorinolarüngoloogia andmete kirjeldus	20
2.4 Pediaatria andmete kirjeldus	21
2.5 Sisehaiguste andmete kirjeldus	22
2.6 Üld- ja plastikakirurgia andmete kirjeldus	24
2.7 Ravi rahastamise lepingu ravijuhtude andmestike kirjeldus	25
3 Mudelite hindamine ja sobivuse analüüs	27
3.1 Otorinolarüngoloogia mudelid	27
3.2 Pediaatria mudelid	30
3.3 Sisehaiguste mudelid	34
3.4 Üld- ja plastikakirurgia mudelid	39
4 Parimate mudelite valimine ja prognoosid	43
4.1 Otorinolarüngoloogia parimad mudelid ja prognoosid	43
4.2 Pediaatria parimad mudelid ja prognoosid	45
4.3 Sisehaiguste parimad mudelid ja prognoosid	47
4.4 Üld- ja plastikakirurgia parimad mudelid ja prognoosid	49
Kokkuvõte	51
Viidatud allikad	53
Lisad	57
Lisa 1 Katuserialade seosed arstlike erialade ja kliinikumi allüksustega . .	57
Lisa 2 Täiendavaid mudelite jooniseid	58
Lisa 2.1 Otorinolarüngoloogia täiendavaid mudelite jooniseid	58
Lisa 2.2 Pediaatria täiendavaid mudelite jooniseid	59
Lisa 2.3 Sisehaiguste täiendavaid mudelite jooniseid	62

Lisa 2.4	Üld- ja plastikakirurgia täiendavaid mudelite jooniseid . . .	64
Lisa 3	Mudelite täpsusnäitajad	66
Lisa 4	Lõplike mudelite Ri väljundid	68

Sissejuhatus

SA Tartu Ülikooli Kliinikum (edaspidi kliinikum) on ülikoolihaigla, baas Tartu Ülikooli õppe- ja teadustööle [1] ning ühtlasi teenuste mahult suurim raviasutus Eestis, kus osutatakse teenuseid kõigil meditsiinierialadel nii ambulatoorse, päevaravi kui statsionaarse ravi raames [2]. Kliinikum on ka üks kahest Eestis tegutsevast piirkondlikust haiglast, mis peavad tagama üld- ja kohalike haiglate vastutuspriirkondades võimekuse osutada kõiki vajalikke eriarstiabi teenuseid, kattes seeläbi kõigi Eesti elanike arstiabi vajaduse [3].

Oma missiooni ning eesmärkide [1] täitmiseks on kliinikumi seisukohalt kasulik näha võimalikult täpselt ette tervishoiuteenuste vajadust, st ravitavate haigete hulka ja jaotumist erialade ja ravitüüpide lõikes. See võiks toetada efektiivset ja eesmärgistatud investeeringute, ressursside ning töökorralduse planeerimist haigla kui terviku, kliinikute ning osakondade tasanditel.

Käesolevas töös on uurimise alla võetud igakuine haiglapatsientide arv neljal ravierialal: otorinolarüngoloogia, pediaatria, sisehaigused ning üld- ja plastikakirurgia. Nendel erialadel kliinikumis haiglaravi vajavate haigete arvu kujunemise potentsiaalseid mõjureid on palju (näiteks haigestumus erinevatesse haigustesse, kokkulepped ja suunad tervishoiusüsteemis, ravijuhendid, demograafilised ja ilmastikust tulenevad tegurid jne), nende vahelised seosed on oletatavasti kompleksed ning nende tegelikud tulevikuväärtused teadmata.

Eelnevast lähtudes on töö põhieesmärgiks prognoosida valitud erialade igakuist haiglapatsientide arvu kliinikumis kuni üks aasta ette, vaadeldes peamiselt prognoose, mis tuginevad antud aegrea minevikuväärtustele. Selleks kasutatakse ARIMA tüüpi mudeleid. Täiendavalt analüüsitakse, kas ühemõõtmeliste mudelite abil saadud prognoose on võimalik täpsustada, võttes arvesse Eesti Haigekassa (edaspidi haigekassa) poolt kliinikumilt aastaks tellitavate ravijuhtude arvu aegrida. Selleks kasutatakse lineaarset regressiooni ARIMA tüüpi vigadega.

Töö esimeses peatükis on toodud ülevaade aegridade analüüsi põhimõistetest, töös kasutatavatest mudelitest ning nende valimise, sobitamise ja võrdlemise meetoditest. Peatükis 2 on antud ülevaade haigete arvu kujunemise taustast ning kõigi töös kasutatud andmestike kirjeldav analüüs. Kolmandas peatükis on toodud kokkuvõtte mudelite hindamise protsessist ja sobivate mudelite valikust. Neljandas peatükis on esitatud saadud mudelite täpsuse analüüs, parima ühe- ja mitmemõõtmelise mudeli valik ning prognoosid.

Töö koostamiseks on kasutatud tekstitöötlusprogrammi \LaTeX . Analüüs on läbi viidud rakendustarkvaras R.

1 Metoodika

1.1 Aegridade teooria mõisteid ja omadusi

Mitteformaalselt nimetatakse aegreaks mingi ajas muutuva suuruse (olgu see tähistatud Z) erinevatele järjestatud ajahetkedele vastavate väärtuste kogumit [4].

Aegridade analüüsi põhiliseks eesmärgiks on konstrueerida matemaatilisi mudeleid, mis võimaldaksid kirjeldada sellise kogumi tekkeprotsessi [5]. Mudelist lähtuvalt saab kontrollida selle sobivust vaadeldava aegrea kirjeldamiseks, tuletada arvutuseeskirjad aegrea erinevate komponentide leidmiseks ning tuleviku prognoosimiseks koos veapiiridega [4]. Sellest eesmärgist tulenevalt eeldatakse, et mõõdetud aegrida on mingi juhusliku protsessi realisatsioon [4], [5].

Definitsioon 1. Juhuslike suuruste peret $\{X_t, t \in \mathbf{T}\}$, mille iga liige X_t on juhuslik suurus nimetatakse *juhuslikuks protsessiks*. Hulka \mathbf{T} nimetatakse juhusliku protsessi *indekshulgaks* [5], [6].

Indekshulk \mathbf{T} on aegridade teoorias mõõtmishetkede hulk. Kui see hulk on diskreetne, siis öeldakse, et tegu on *diskreetse ajaga aegreaga*, vastasel juhul *pideva ajaga aegreaga* [4], [5].

Antud töös käsitletakse ainult diskreetse ajaga aegridu, mille korral mõõtmised toimuvad fikseeritud võrdsete ajavahemike tagant. Sellisel juhul saab mõõtmishetked tähistada $\tau_t := \tau_0 + ih$, kus h on fikseeritud mõõtmishetkede vaheline ajavahemik ning $i \in \mathbb{Z}$. Niisiis eeldatakse, et uuritava aegrea ajahetkel $\tau_t, t \in \mathbb{Z}$, mõõdetud väärtus z_t on mingi juhusliku suuruse Z_t realisatsioon.

Suuruse Z muutumisel ajas võib olla erinevaid põhjuseid. Aegridade analüüsimisel eristatakse sageli järgmisi komponente [4], [7]:

- trend ehk pikaajaline kindlasuunaline muutus (kasvamine või kahanemine), kusjuures trend võib olla ajas muutumatu iseloomuga ehk globaalne või ajas pikkamööda muutuv ehk lokaalne;
- perioodilised või sesoonsed muutused ehk kindla perioodiga muutused, mis sõltuvad näiteks kellaajast, nädalapäevast, kuust;
- juhuslik komponent ehk ebaregulaarsed, lühikeste ajavahemike järel toimuvad muutused.

Lisaks eristatakse mõnel juhul ka tsükli komponenti, mis kujutab endast ebaregulaarse pikema perioodiga kindlasuunalisi muudatusi (nt majandustsükkel)

[7], aga selle eraldamisele trendi komponendist käesolevas töös tähelepanu ei pöörata.

Järgnevas tuginetakse allikatele [4], [5].

Üks oluline aegridu kirjeldavate juhuslike protsesside klass on statsionaarsed juhuslikud protsessid. Statsionaarsuse eeldusele tuginevad mitmed aegride mudelid, sh antud töös kasutatavad mudelid.

Definitsioon 2. Juhuslikku protsessi $Z_t, t \in \mathbb{Z}$ nimetatakse (*tugevalt*) *statsionaarseks*, kui iga täisarvude komplekti t_1, \dots, t_m ja iga täisarvu q korral on juhuslikud vektorid $(Z_{t_1}, \dots, Z_{t_m})$ ja $(Z_{t_1+q}, \dots, Z_{t_m+q})$ sama jaotusega.

Definitsioonis 2 antud statsionaarsuse eeldus on enamasti rakendamiseks liiga tugev ning praktikas raskesti kontrollitav. Seetõttu kasutatakse sageli nõrga statsionaarsuse mõistet, mis seab teatavad piirangud ainult juhusliku protsessi (esimestele) momentidele.

Definitsioon 3. Juhuslikku protsessi $Z_t, t \in \mathbb{Z}$ nimetatakse *k-ndat järku nõrgalt statsionaarseks*, kui iga täisarvude komplekti t_1, \dots, t_m ja iga täisarvu q korral on juhuslike vektorite $(Z_{t_1}, \dots, Z_{t_m})$ ja $(Z_{t_1+q}, \dots, Z_{t_m+q})$ kõik kuni *k*-ndat järku momendid võrdsed.

Statsionaarne protsess Z_t on *k-ndat järku nõrgalt statsionaarne*, kui $E(Z_t^k)$ on lõplik.

Edaspidi kasutatakse antud töös mõisteid *statsionaarne* ja *nõrgalt statsionaarne* tähenduses teist järku nõrgalt statsionaarne.

Aegrea erinevate hetkede vaatlused ei ole reeglina sõltumatud. Lineaarset sõltuvust aegrea kahele erinevale ajahetkele vastavate juhuslike suuruste vahel mõõdab autokovariatsioon ehk kovariatsioon vaadeldava protsessi nendele ajahetkedele vastavate juhuslike suuruste vahel.

Olgu $Z_t, t \in \mathbb{Z}$ lõpliku dispersiooniga.

Definitsioon 4. *Autokovariatsioonifunktsioon* defineeritakse järgmiselt:

$$\gamma(s, t) = \text{Cov}(Z_s, Z_t) = E[(Z_s - E(Z_s))(Z_t - E(Z_t))].$$

Kuna mugavam on sõltuvust kirjeldada suurusega lõigus $[-1, 1]$, siis enamasti kasutatakse aegrea kahele erinevale ajahetkele vastava juhusliku suuruse vahelise sõltuvuse kirjeldamiseks autokorrelatsioonifunktsiooni.

Definitsioon 5. Autokorrelatsioonifunktsioon (ACF) on defineeritud järgmiselt:

$$\rho(s, t) = \text{Cor}(Z_s, Z_t) = \frac{\gamma(s, t)}{\gamma(s, s)\gamma(t, t)}.$$

Analoogselt klassikalise korrelatsiooniga on autokorrelatsioon 1 või -1 , kui nende suuruste vahel esineb funktsionaalne lineaarne seos, st $Z_t = \beta_0 + \beta_1 Z_s$.

Definitsioonist 3 järeldub, et teist järku nõrgalt statsionaarse protsessi Z_t korral on

$$\begin{aligned} E(Z_t) &= \mu, \\ D(Z_t) &= \sigma^2, \end{aligned}$$

kus μ ja σ on konstandid, ning autokovariatsioonifunktsioon $\gamma(t, t + p)$ sõltub ainult ajamomentide vahest p . Seega saab autokovariatsiooni- ja autokorrelatsioonifunktsiooni iga $p \in \mathbb{Z}$ kirja panna järgmiselt:

$$\begin{aligned} \gamma(p) &:= \gamma(t, t + p) = \text{Cov}(Z_t, Z_{t+p}), \\ \rho(p) &:= \rho(t, t + p) = \text{Cor}(Z_t, Z_{t+p}) = \frac{\gamma(p)}{\sigma^2}. \end{aligned}$$

Lisaks autokorrelatsioonidele räägitakse aegridade teoorias sageli ka osaaautokorrelatsioonidest. Osaaautokorrelatsiooni mõiste tugineb osakorrelatsiooni mõistele (vt täpsemalt nt [4]).

Olgu $\tilde{Z}_t = Z_t - E(Z_t)$.

Definitsioon 6. Teist järku nõrgalt statsionaarse protsessi Z k -ndat järku osaaautokorrelatsioonikordajaks nimetatakse suuruste \tilde{Z}_t ja \tilde{Z}_{t-k} osakorrelatsiooni pärast suuruste $\tilde{Z}_{t-1}, \dots, \tilde{Z}_{t-(k-1)}$ mõju eemaldamist.

Praktikas on kasutada ainult teatav hulk vaadeldud aegrea väärtusi $z_t, t = 1, \dots, T$, mille põhjal on võimalik arvutada empiirilist keskmist, autokovariatsiooni- ja autokorrelatsioonifunktsiooni järgmiselt:

$$\begin{aligned} \bar{\mu} &= \frac{1}{T} \sum_{t=1}^T z_t; \\ c_p &= \frac{1}{T} \sum_{t=1}^{T-p} (z_t - \bar{\mu})(z_{t+p} - \bar{\mu}), p = 0, 1, \dots, T-1; \\ r_p &= \frac{c_p}{c_0}. \end{aligned}$$

Tuleb tähele panna, et kui protsess ei ole vähemalt nõrgalt statsionaarne, ei ole saadud empiirilised näitajad seotud vastavate mõistete teoreetilise sisuga.

1.2 Ühemõõtmelised mudelid

Käesolev punkt põhineb allikal [4], kui pole viidatud teisiti.

Antud töös otsitakse aegridadele ühemõõtmelisi mudeleid ARIMA tüüpi mudelite ehk integreeritud autoregressiivsete liikuva keskmisega (*Autoregressive Integrated Moving Average*) mudelite klassist.

Definitsioon 7. Aegrea mudelit nimetatakse *ARIMA tüüpi mudeliks*, kui Z_t avaldub lõpliku arvu varasemate $Z_i, i < t$, ja lõpliku arvu sõltumatute sama jaotusega häirituste $A_i, i \leq t$ lineaarkombinatsioonina kujul

$$Z_t = \phi_0 + \sum_{i=1}^p \phi_i Z_{t-i} + A_t + \sum_{i=1}^q \psi_i A_{t-i}.$$

Järgnevas kasutatakse kirjapaneku lihtsustamiseks nihkeoperaatorit B , mis on defineeritud võrdusega $BZ_t = Z_{t-1}, \forall t$. Olgu defineeritud ka järgmised funktsioonid:

$$\phi(x) = 1 - \sum_{i=1}^p \phi_i x^i, \quad (1)$$

$$\theta(x) = 1 - \sum_{i=1}^q \theta_i x^i, \quad (2)$$

kus $\phi_i, i = 1, \dots, p$, ja $\theta_i, i = 1, \dots, q$, on reaalarvulised kordajad.

Definitsioon 8. Olgu $A_t, t \in \mathbb{Z}$ tsentreeritud, mittekorreleeritud ja teist järku nõrgalt statsionaarne protsess. Omagu protsess $Z_t, t \in \mathbb{Z}$ konstantset keskvaartust μ ning tähistagu \tilde{Z}_t protsessile Z_t vastavat tsentreeritud protsessi, $\tilde{Z}_t = Z_t - \mu$.

- *Autoregressiivseteks protsessideks järguga p ehk $AR(p)$ protsessideks* nimetatakse teist järku nõrgalt statsionaarseid protsesse kujul

$$\tilde{Z}_t = \sum_{i=1}^p \phi_i \tilde{Z}_{t-i} + A_t$$

ehk $\phi(B)\tilde{Z}_t = A_t$.

- *Liikuva keskmise protsessideks järguga q ehk $MA(q)$ protsessideks* nimetatakse

protsesse kujul

$$\tilde{Z}_t = A_t - \sum_{i=1}^q \theta_i A_{t-i}$$

ehk $\tilde{Z}_t = \theta(B)A_t$.

- *ARMA(p, q) protsessideks* nimetatakse teist järku nõrgalt statsionaarseid protsesse kujul

$$\tilde{Z}_t = \sum_{i=1}^p \phi_i \tilde{Z}_{t-i} + A_t - \sum_{i=1}^q \theta_i A_{t-i}$$

ehk $\phi(B)\tilde{Z}_t = \theta(B)A_t$.

Tihti pole täidetud aegrea statsionaarsuse nõue, kuna keskmine on ajas muutuv. Siinkohal tuleb taustainfole tuginedes otsustada, kas keskmise muutumine ajas on seotud mingi globaalse deterministliku trendiga (näiteks kui aegrea tõusev tase on seotud mingi sihipärase tegevusega, mis vaadeldava suuruse väärtust ajas edasi liikudes tõstab) või pigem stohhastilise trendiga, mis võib aja jooksul muutuda. Kui on alust eeldada globaalset trendi, siis võib selle aegreast eemaldada, hinnates sobiva ajast sõltuva funktsiooniga trendikõvera ning edasises otsida jääkidele sobivat ARMA mudelit (kui saadud jääkaegrida on statsionaarne). Kui pole selget põhjust globaalse trendi eeldamiseks, aga aegrea muudud (või muutude muudud) vastavad statsionaarsuse nõuetele, võivad sobida nn integreeritud ARMA mudelid ehk ARIMA mudelid.

Definitsioon 9. *ARIMA(p, d, q) protsessideks* nimetatakse juhuslikke protsesse Z_t , mille d -ndat järku muudud $W_t = (1 - B)^d Z_t$ vastavad $ARMA(p, q)$ protsessile. St W_t on teist järku nõrgalt statsionaarne ning rahuldab

$$\phi(B)\tilde{W}_t = \theta(B)A_t,$$

kus $\tilde{W}_t = W_t - E(W_t)$, juhuslikud suurused A_t on tsentreeritud, sama dispersiooniga ja mittekorreleeritud ning funktsioonid ϕ ja θ on defineeritud nagu (1) ja (2).

Diferentsimise vajadusele viitab selge aegrea taseme muutumine ajas, mida kinnitab autokorrelatsioonide aeglane kahanemine [8].

Sageli pole aegrea statsionaarsuse eeldus täidetud aegreas peituva sesoonse komponendi tõttu. Sellisel juhul on võimalik püüda sesoonsust arvesse võtta ja/või kasutada sesoonset diferentsimist, et saavutada statsionaarsus.

Olgu defineeritud funktsioonid

$$\Phi(x) = 1 - \sum_{i=1}^P \Phi_i x^i, \quad (3)$$

$$\Theta(x) = 1 - \sum_{i=1}^Q \Theta_i x^i, \quad (4)$$

kus $\Phi_i, i = 1, \dots, P$, ja $\Theta_i, i = 1, \dots, Q$, on reaalarvulised kordajad.

Definitsioon 10. *Perioodiga s multiplikatiivseteks sesoonseteks $ARIMA(p, d, q)(P, D, Q)_s$ tüüpi mudeliteks, lühidalt ka $SARIMA(p, d, q)(P, D, Q)_s$ tüüpi mudeliteks* nimetatakse mudeleid kujul

$$\phi(B)\Phi(B^s)\tilde{W}_t = \theta(B)\Theta(B^s)A_t,$$

kus $\tilde{W}_t = W_t - E(W_t)$ ja $W_t = (1 - B)^d(1 - B^s)^D Z_t$ ning funktsioonid $\phi, \theta, \Phi, \Theta$ on defineeritud nagu (1)–(4),

Sesoonsust võib püüda arvesse võtta sesoonse(te) parameetri(te) lisamisega ehk sobiva P või Q määramisega. Kui sesoonsus on tugev ja stabiilne, millele viitab näiteks sesoonsete autokorrelatsioonide aeglane kahanemine, siis võib olla põhjust rida sesoonselt diferentsida [8].

Statsionaarsele aegreale (või aegrea teisendusele, mida saab lugeda statsionaarseks) sobitatavate $ARMA(p, q)$ mudelite valimisel tuginetakse allika [8] poolt kirjeldatud omadustele, mis antud protsesside teoreetiliste autokorrelatsioonide ja osaaautokorrelatsioonide kohta kehtivad. Lühülevaade nendest omadustest on järgmine.

- Osaaautokorrelatsioonide võrdumine nulliga alates järgust $p + 1$ ning autokorrelatsioonide järkjärguline vähenemine on omased autoregressiivsele protsessile $AR(p)$.
- Autokorrelatsioonide võrdumine nulliga alates järgust $q + 1$ ning osaaautokorrelatsioonide järkjärguline kahanemine on omane liikuva keskmise protsessile $MA(q)$.
- $ARMA(p, q)$ -protsessile on omane autokorrelatsioonide ja osaaautokorrelatsioonide järkjärguline vähenemine. Antud töös on vaadeldud vaid $ARMA(1, 1)$ -protsessi, mille korral autokorrelatsioonid kahanevad eksponentsiaalselt pärast järku 1 ning osaaautokorrelatsioonid pärast järku 1 käituvad sarnaselt $MA(1)$ protsessile.

Otsustamaks, millise protsessiga tegu võiks olla, uuritakse nende omaduste kehtivust konkreetse aegrea korral selle empiiriliste autokorrelatsioonide ja osaaautokorrelatsioonide jooniste abil.

Sõltumatutele sama jaotusega juhuslikele suurustele vastava aegrea korral on suurused $r_p, p > 0$ asümptootiliselt normaaljaotusega, keskvärtusega ligikaudu $-\frac{1}{T}$ ja standardhälbega $\frac{1}{\sqrt{T}}$. Järelikult peaks piisavalt suure T korral iga $p > 0$ korral r_p jääma tõenäosusega 0,95 vahemikku $\left[\frac{-1}{T} - \frac{\Phi^{-1}(0,975)}{\sqrt{T}}, \frac{-1}{T} + \frac{\Phi^{-1}(0,975)}{\sqrt{T}} \right]$, kus Φ on standardse normaaljaotuse jaotusfunktsioon. Joonisele, mille põhjal mudelit valitakse, kantaksegi autokorrelatsioonid ning vastavad piirid. Analoogselt tehakse osaaautokorrelatsioonidega.

Sesoonse käitumise olemasolul kasutatakse samalaadset mudelite identifitseerimise viisi, leides võimalikud mudelid eraldi $\text{ARIMA}(p, d, q)(P, D, Q)$ sesoonsele ja mittesoonsele osale. Seejuures vaadeldakse P ja Q välja selgitamisel sesoonseid ehk perioodi s kordsetele ($is, i = 1, 2, \dots$) järkudele vastavaid autokorrelatsioone ja osaaautokorrelatsioone. Mittesoonse osa mudeli välja selgitamiseks vaadeldakse sesoonsuse perioodi siseseid autokorrelatsioone ja osaaautokorrelatsioone.

Üldiselt eelistatakse töös võimalusel $\text{AR}(p)$ ja $\text{MA}(q)$ mudelit, kui saab vältida segamudelit, kuna viimast on keerulisem interpreteerida.

Käesolevas töös vaadeldavate aegridade puhul on tegu loendusandmetega, mis tähendab, et aegrea võimalikud väärtused on mittenegatiivsed täisarvud. Teoorias ei ole sellisel juhul täidetud normaaljaotuse eeldus. Sellest lähtuvalt kasutatakse kõigi mudelite hindamisel tinglike ühesammuliste prognooside vigade ruutude summa ehk CSS (*Conditional Sum of Squares*) minimiseerimise meetodit, mitte sageli kasutatavat tõepära maksimiseerimise meetodit (mis eeldab aegrea moodustavate juhuslike suuruste normaaljaotust). Tinglik tähendab siinjuures seda, et esimeste prognooside jaoks vajalikud puuduvad väärtused asendatakse keskvärtustega.

1.3 Mitmemõõtmelised mudelid

Antud alapunktis tuginetakse allikale [4].

Üks võimalus huvipakkuva aegrea tulevikuväärtuste prognoosimiseks, kasutades lisaks ka teiste aegridade andmeid, on mitmene lineaarne regressioon ARIMA tüüpi vigadega ehk ARIMAX -mudel.

Olgu Z_t huvipakkuva tunnuse väärtus ajal t ning $(X_1(t), \dots, X_m(t))$ suuruse Z_t prognoosimiseks kasutatav argumenttunnuste vektor.

Definitsioon 11. *Mitmeseks lineaarseks regressioonimudeliks ARIMA tüüpi vigadega ehk ARIMAX-mudeliks nimetatakse mudelit kujul*

$$Z_t = \beta_0 + \sum_{i=1}^m \beta_i X_i(t) + \varepsilon_t,$$

kus vead ε_t vastavad mingile ARIMA tüüpi protsessile.

ARIMAX-mudel sobitatakse käesolevas töös järgmiselt.

1. Esmalt sobitatakse andmete le mitmene regressioonimudel vähimruutude meetodil.
2. Seejärel analüüsitakse saadud mudeli vigade käitumist ning valitakse ARIMA mudeli kuju suuruste ε_t jaoks.
3. Lõpuks leitakse valitud ARIMAX tüüpi mudeli parameetrid CSS-meetodil.

Antud töös huvi pakkuvad regressorread on aastase intervalliga, aga uuritav rida on kuu tasandil. Et aasta tasandil regressoreid kuu tasandil aegridade mudelisse kaasata, teisendati regressorread kuu tasandile. Selleks loodi uued kuu täpsusega tükiti konstantsed aegread järgmiselt: kui x_t tähistab algse regressorrea väärtust aastal t , siis uues aegreas loeti aasta t kuu i väärtuseks $y_{i(t)} = \frac{1}{12}x_t$.

Sellisel kujul aegridade kasutamine regressoritena on mitmes mõttes problemaatiline. Põhiline probleem seisneb selles, et regressorreas toimub taseme muutus järsult iga aasta alguses. Selle tulemusel on oodata, et regressioonimudeli jäägid on iga aasta esimesel kuul suured. See tähendab, et jääkide aegrida võib sisaldada kunstlikult tekitatud sesoonsust. Samas ei ole antud teema kontekstis lihtne leida ka muud, n-ö siledamat aegrea kuu tasandile teisendamise viisi, mis oleks selgelt põhjendatav.

Alternatiivne võimalus oleks analüüsida seoseid aasta tasandil (agregeerides kuu tasandil olevad read), ent siis on probleemiks selle seose rakendamine kuu tasandil prognooside täpsustamiseks. Antud juhul oleks aasta tasandil andmete analüüsimine ka teistpidi veaohalik, kuna vaadeldavate regressorriade pikkus on kõigest 8 aastat.

1.4 Mudelite sobivuse ja täpsuse analüüs

Antud töös jagatakse iga huvipakkuv aegrida kaheks osaks: treening- ja testandmeteks. Täpsemalt, olgu kõik aegrea mõõdetud väärtused $z_t, t = 1, \dots, T$ jagatud kaheks:

- treeningandmed, s.o teadaolevad väärtused z_t ajahetkeni $k = T - n$, ja
- testandmed ehk viimased n teadaolevat väärtust $z_t, t = k + 1, \dots, T$.

Antud töös kasutatavad aegread ei ole eriti pikad, mistõttu jäetakse testandmestikku vaid ühe aasta andmed, st $n = 12$. Mudelid hinnatakse ainult treeningandmeid kasutades. Mudelite täpsuse võrdlemiseks kasutatakse kõiki andmeid.

Hindamaks mudeli sobivust andmetega kasutatakse mudeli jääke $e_t = z_t - \hat{z}_{t|t-1}, t = 1, \dots, k$. Aegridade mudelit võib lugeda andmetega sobivaks, kui mudeli jäägid ei ole süstemaatilised ehk need ei sisalda enam informatsiooni, mida saaks kasutada prognooside parandamiseks [4]. Põhiline sobivuse kriteerium, mida mudel peaks täitma, on jääkide sõltumatus [4].

Sõltumatuse uurimiseks vaadeldakse jääkide autokorrelatsioone, mis sõltumatute jääkide korral peaks olema nullid [4]. Selleks uuritakse jääkide autokorrelatsioonide joonist, kuhu on kantud ka piirid, nagu on kirjeldatud punktis 1.2. Lõplik hinnang mudeli jääkide sõltumatuse kohta antakse teatava suurusega grupi autokorrelatsioonide nullist erinevuse testi põhjal.

Olgu vaatluse all autokorrelatsioonid r_1, \dots, r_m . Ljung-Box testi teststatistik arvutatakse järgmiselt [4], [9]:

$$Q = T(T+2) \sum_{p=1}^m \frac{r_p^2}{T-p}.$$

Vaadeldava autokorrelatsioonide grupi suuruse m valikul lähtutakse töös allikast [7], mis soovib sesoonsete aegridade korral kasutada vähemalt $h = 2s$, kus s vastab sesoonsuse perioodile. See tähendab, et mudel loetakse sobivaks, kui vähemalt grupi $h = 2s$ korral Ljung-Box test ei lükka ümber nullhüpoteesi, et autokorrelatsioonid kuni vaheni h on võrdsed nulliga.

Statistik Q on sõltumatutele sama jaotusega juhuslikele suurustele vastava aegrea puhul asümptootiliselt χ^2 -jaotusega vabadusastmetega m [4], [9]. Mudeli jääkide sõltumatuse uurimisel Ljung-Box testi abil soovitakse arvesse võtta mudelis hinnatud parameetrite arvu M [4], [7], mida on ka käesolevas töös tehtud. See tähendab, et mudeli jääkide sõltumatuse testimisel eeldatakse statistiku Q kohta, et see on χ^2 -jaotusega vabadusastmetega $m - M$ [7].

Mudelite täpsuse hindamiseks ja erinevate (ka ühe- ja mitmemõõtmeliste) mudelite võrdlemiseks kasutatakse töös täpsusnäitajaid, mille arvutamisel kõrvutatakse mudeli abil leitud prognoose ja tegelikke väärtusi. Kasutatav täpsusnäitaja on ruutkeskmine viga (*Root Mean Square Error* ehk RMSE).

Kuna testandmete periood on kõigest 12 kuud ehk 12-sammulise prognoosi täpsuse hindamiseks on vaid üks katse, siis ei ole ainult testperioodi RMSE kasutamine mudelite võrdlusel mõistlik. Kuna aastase prognoosi puhul sõltub teise väärtuse

prognoosimisel tehtav viga esimese väärtuse prognoosiveast jne, siis võib aastase prognoosi $RMSE$ olla küllalt suure dispersiooniga juhuslik suurus ja seetõttu ei saa ühe katse põhjal kuigi head hinnangut selle tegeliku keskmise väärtuse kohta. Selleks, et proovida saada paremat ülevaadet mudeli täpsusest aasta jagu ette prognoosimisel, uuritakse 12-sammuliste prognooside ruutkeskmisi vigu eraldi ka treeningperioodis.

Esiteks leitakse treeningperioodi aastate prognooside ruutkeskmised vead ning nende koondnäitaja järgmiselt.

1. Kasutades hinnatud mudeli järgi fikseeritud parameetritega mudelit ning kõigi eelnevate aastate andmeid, prognoositakse iga järgmise kalendriaasta 12 kuu väärtused. Seejuures kasutatakse vähemalt prognooside leidmiseks minimaalselt vajalikku hulka andmeid täisaastates (konkreetne vajalik aastate arv sõltub mudelist). Näiteks kui vaadeldava mudeli abil prognoosimiseks on vaja teada aegrea eelneva 13 kuu väärtusi, siis esimene aasta, millele prognoose leitakse, on vaadeldava aegrea kolmas aasta ning prognooside leidmiseks kasutatakse kahe täisaasta andmeid; järgmise aasta jaoks kolme täisaasta andmeid jne. Selliselt saadakse iga võimaliku aasta $j = j_1, \dots, j_A$ kohta iga kuu prognoosid $\hat{z}_{t|i_j}, t = i_j + 1, \dots, i_j + 12$ ning vastavad prognoosivead

$$e_{t|i_j} = z_t - \hat{z}_{t|i_j}, t = i_j + 1, \dots, i_j + 12,$$

kus i_j on aastale $j - 1$ vastav detsembrikuu.

2. Seejärel arvutatakse iga võimaliku aasta $j = j_1, \dots, j_A$ jaoks 12 kuu prognooside ruutkeskmise viga:

$$RMSE_j = \sqrt{\frac{1}{12} \sum_{t=i_j+1}^{i_j+12} e_{t|i_j}^2},$$

kus i_j on aastale $j - 1$ vastav detsembrikuu. Antud näitaja iseloomustab 12 kuu prognoosi täpsust vaadeldaval aastal.

3. Lisaks leitakse treeningperioodi aastate $RMSE$ -de põhjal ka koondnäitaja, nn treeningaastate ruutkeskmise $RMSE$ järgmise valemi alusel:

$$\overline{RMSE} = \sqrt{\frac{1}{A} \sum_{j=j_1}^{j_A} RMSE_j^2}.$$

Antud näitaja põhjal võiks olla võimalik võrrelda mudelite stabiilsust aastaste

prognooside tegemisel treeningperioodis.

Teiseks leitakse testperioodi 12 kuu prognoosi RMSE. Kasutades hinnatud mudelit ja kõiki treeningandmeid, leitakse testaasta 12 kuu prognoosid ning arvutatakse vead $e_{t|k} = z_t - \hat{z}_{t|k}, t = k + 1, \dots, T$. Nende põhjal leitakse RMSE:

$$RMSE = \sqrt{\frac{1}{12} \sum_{t=k+1}^T e_{t|k}^2}.$$

Kuna 12-sammulise prognoosi arvutamisel kasutatakse kõiki vahepealseid prognoose, st saadud prognoosid on üksteisest sõltuvad, vaadeldakse testperioodis lisaks ka ühesammuliste prognooside ruutkeskmist viga. See tähendab, et iga järgmise prognoosi arvutamiseks kaasatakse eelmis(t)ele ajahetke(de)le vastav(ad) tegelik(ud) väärtus(ed). Seejärel leitakse vead $e_{t|t-1} = z_t - \hat{z}_{t|t-1}, t = k + 1, \dots, T$, ning ruutkeskmine viga:

$$RMSE^* = \sqrt{\frac{1}{12} \sum_{t=k+1}^T e_{t|t-1}^2}.$$

Täiendavalt tuuakse mudelite võrdluses välja ka mudeli standardviga ehk mudeli ühesammuliste prognooside vigade standardhälbe hinnang $\hat{\sigma}$.

Parim ühe- ja mitmemõõtmeline mudel selgitatakse välja eelnevalt toodud täpsusnäitajate võrdlemisel, kusjuures sarnase täpsusega mudelite korral eelistatakse lihtsamat mudelit ehk mudelit, milles on hinnatud vähem parameetreid või millel on muus mõttes lihtsam kuju (nt sisaldab ainult AR või MA liikmeid).

Parima ühe- ja mitmemõõtmelise mudeli jaoks antakse 12-sammuline prognoos testaasta kohta koos 95%-prognoosiintervallidega, mis arvutatakse kui

$$\hat{z}_{T+h|T} \pm \Phi^{-1}(0,975)\hat{\sigma}_h,$$

kus Φ on standardse normaaljaotuse jaotusfunktsioon ja σ_h on h -sammulise prognoosi jaotuse standardhälbe hinnang [7]. Prognoosiintervall on korrektne, kui on täidetud häirituste normaaljaotuse eeldus (ning mudeli kordajad on õiged).

2 Andmete kirjeldus

Andmete kirjeldamisel kasutatakse järgnevaid ravitöö ja raviarveldusega seotud mõisteid (kohandatud allikatest [10], [11], [12]).

- *Statsionaarne arstiabi* – tervishoiuteenused, mille osutamiseks on vajalik inimese ööpäevaringne viibimine haiglas.
- *Päevaravi* – tervishoiuteenused, mille korral patsient vajab ravi või uuringute tõttu jälgimist mõne tunni jooksul ravivoodis, aga lahkub ööseks koju.
- *Vältimatu arstiabi* – tervishoiuteenus, mida osutatakse olukorras, kus abi edasilükkamine või selle andmata jätmine võib põhjustada abivajaja surma või püsiva tervisekahjustuse.
- *Statsionaarne haigusjuht* – patsiendi viibimine haiglas statsionaarse arstiabi saamise eesmärgil haiglasse saabumisest haiglast lahkumise (koju / teise haiglasse) või surmani.
- *Statsionaarne ravijuht* (ravi rahastamise leping) – raviarve, millel kajastuvad ühele ravikindlustatud isikule kogu statsionaarse haigusjuhu käigus tehtud terviseuuringud ja osutatud teenused.

2.1 Haigete arvu kujunemine

Eesti tervishoiusüsteem on üles ehitatud solidaarsele ravikindlustusele: töölkäivate inimeste palgast makstud sotsiaalmaksust koguneva rahaga tagatakse juurdepääs arstiabile kõigi ravikindlustatute jaoks, kusjuures ravikindlustatute alla kuuluvad lisaks töötavatele inimestele ka näiteks alaealised, tudengid, pensionärid jpt grupid [13]. Riikliku ravikindlustuse haldaja on haigekassa ning tervishoiuteenuseid rahastataksegi peamiselt riigieelarvest haigekassa vahendusel [13]. Ka enamik (94%) kliinikumis osutatud tervishoiuteenustest on rahastatud haigekassa poolt vastavalt ravi rahastamise lepingule (edaspidi RRL) [2]. Selleks sõlmitakse iga aasta algul lepingu lisa, milles lepitakse kokku haigekassa poolt kaetav ravijuhtude maksumus ja vastav minimaalne ravijuhtude arv [14].

Tervishoiuteenuste osutajatelt haigekassa poolt ostetavate ravijuhtude hulga planeerimisel hinnatakse tervishoiuteenuse nõudlust ning võetakse arvesse teenuste geograafilist, ajalist kättesaadavust jm [15]. Nõudlus, mis tähendab kindlustatute terviseseisundist tulenevat vajadust tervishoiuteenuste kasutamiseks, sõltub näiteks kindlustatute vanuselise ja soolise koosseisu muutustest, meditsiinitehnoloogiarengust ning tervishoiukorraldusest [15], [16].

Võib arvata, et lepingumahul on mõju aastas ravitavate patsientide arvule, vähemalt selles osas, mis puudutab plaanilist (mitte vältimatut) ravi. Ühest küljest on teenuse osutajal huvi täita kokkulepitud mahtu, teisest küljest katab üle piirmahu osutatud teenuste maksumusest haigekassa vaid osa¹.

Töös vaatluse all olevas haigete arvus kajastub ka vältimatu arstiabi, mida on tervishoiuteenuste osutaja kohustatud andma kõigile vajajatele. Vältimatu ravi maht, mis moodustab märkimisväärse osa kliinikumis osutatavast (statsionaarsest) ravist [14], ei tohiks seega olla otseselt mõjutatud haigekassaga sõlmitavast RRList, vaid enam nendest teguritest, millele vastavalt kujuneb tervishoiuteenuste nõudlus üldiselt. Lisaks võib eeldada, et eriti vältimatu arstiabi vajadust ühes kuus võivad mõjutada ka sesoonsed näitajad, nt ilm või aastaajale omased tegevused.

Võib arvata, et lisaks üldisele tervishoiuteenuste nõudlusele, haigestumustrendidele ning eelnevalt mainitud potentsiaalsetele mõjuritele, mõjutavad ravitud haigete koguarvu aastas ning nende jagunemist kuude vahel ka ruumide, voodikohtade ja muude ressursside olemasolu ja kasutamise töökorraldus, personali olemasolu (sh personali haigestumine), puhkustega seotud töökorraldus (seda ilmselt rohkem plaanilise töö osas) jne.

Kokkuvõttes on potentsiaalseid mõjureid konkreetse ravieriala haigete arvu kujunemisel kuus palju ning nende omavahelised seosed arvatavasti keerulised.

2.2 Haigete arvu andmestike moodustamine ja korrastamine

Kliinikum jaguneb 17 kliinikuks, mis koondavad enda alla ühe või mitme seotud eriala pädevused [2]. Kliinikud koosnevad erinevatest osakondadest ja allüksustest, mis tegelevad kitsamate erialadega erinevate ravitüüpide raames. Kliinikumi täpse struktuuri määrab kliinikumi juhtimissüsteemi dokument TKL-01 [17], kus on määratletud ka seosed struktuuriüksuste ning ravi rahastamise lepingu profiilide ja põhierialadega. Tuginedes selle dokumendi kehtivale ning varem kehtinud versioonidele võeti antud töös vaatluse alla neli n-ö katuseriala, mille alla kaasati sobivad allüksused ning mis on edasises nimetatud järgmiselt:

- (a) otorinolarüngoloogia (LOR),
- (b) pediatría,
- (c) sisehaigused,
- (d) üld- ja plastikakirurgia (ÜPK).

¹See ei kehti teatud erijuhtudele, mille korral haigekassa tasub kogu kulu vastavalt nende esinemisele, nt sünnitused, organsiirdamised jm. Vt lähemalt [12], [14].

Nende katuserialade all vaadeldav ravitegevus ei vasta üksüheselt samanimeliste arstlike erialade tegevusele, vaid võib hõlmata ainult osa selle eriala tegevusest või hoopis mitut eriala. Katuseriala alla kaasatud arstlikest erialadest on ülevaade antud alapunktides 2.3–2.6 ning seoste kirjeldus nii erialade kui kliinikumi struktuuriüksustega on antud lisas 1, tabelis L1.1 (lk 57).

Erialal (*statsionaarselt*) *ravitud haigete arv* ehk *haiglapatsientide arv* tähendab antud töö kontekstis statsionaarsete haigusjuhtude arvu, mille vältel patsienti raviti mõnes selle eriala alla arvatud allüksuses. Haige kajastub vastavalt haigusjuhu lõppemise kuule iga allüksuse all, kus ta ravi jooksul viibis, ühekordselt (ent võib mõnel juhul olla ühe eriala all mitmekordselt, kui patsient viibis mitmes selle eriala alla kuulavas üksuses sama haigusjuhu jooksul).

Antud töös on kasutatud ainult kliinikumi elektroonsest andmebaasist kättesaadavaid haigete arvu andmeid. Andmed päriti kuude ja allüksuste tasandile agregeeritud kujul kogu kättesaadava perioodi ulatuses ehk jaanuarist 2008 kuni detsembrini 2018. Seejärel teostati topeltloendamise kontroll struktuurimuudatustele järgneva kuu kohta ning viidi sisse vajalikud parandused: kui üks ja sama haigusjuht sattus sellel kuul nii suletud üksuse kui selle järglase arvestusse, siis kajastati teda ühekordselt uue üksuse real. Lõpuks agregeeriti saadud andmed ülnimetatud erialadeks vastavalt tabelis L1.1 kirjeldatud seostele.

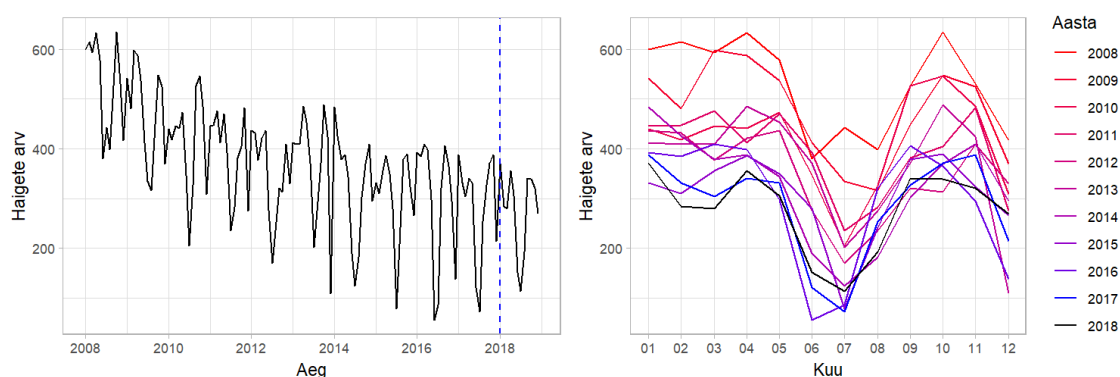
Oli alust arvata, et seoses järkjärgulise üleminekuga elektroonsele haiguslugude täitmisele aastatel 2008–2010 võivad saadud aegridade alguskuudel esineda mõningad probleemid või erinevused dokumenteerimises. Seetõttu võrreldi andmete käitumist aegrea esimesel 6 kuul ülejäänud perioodiga, et hinnata üleminekuaja andmete sobivust töös kasutamiseks. Ühegi aegrea alguses ei esinenud märkimisväärsed kõrvalekaldeid selle järgeva perioodi keskmisest käitumisest, seega võis kõik read lugeda täies mahus sobivaks.

Igast haigete arvu andmestikust moodustati treeningandmestikud ja testandmestik. Ühemõõtmeliste mudelite hindamiseks kasutatavasse treeningandmestikku võeti kõik vaatlused aegrea algusest kuni detsembrini 2017. Mitmemõõtmeliste mudelite hindamisel kasutatav treeningandmestik sisaldas vaatlusi jaanuarist 2011 kuni detsembrini 2017, kuna regressorriidade andmed olid olemas alates aastast 2011. Testandmestikku võeti viimased 12 vaatlust ehk 2018. aasta andmed.

2.3 Otorinolarüngoloogia andmete kirjeldus

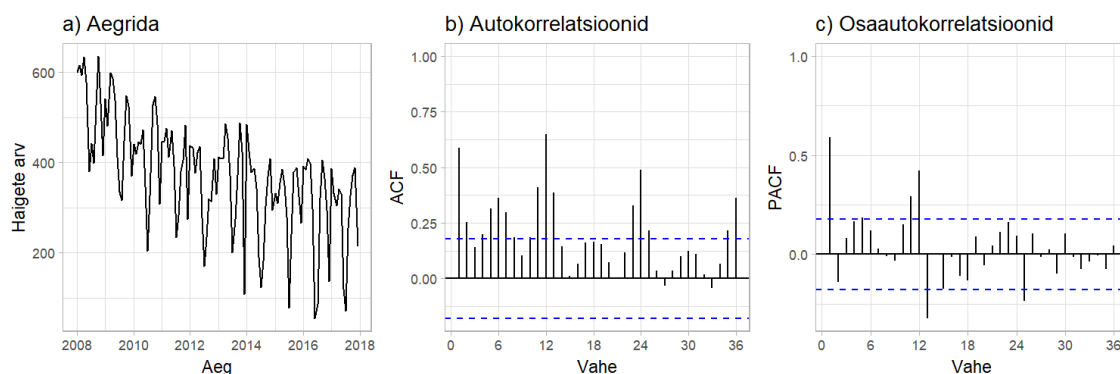
Otorinolarüngoloogia (edaspidi LOR) hõlmab kõrva-nina-kurguhaiguste diagnostikat ja ravi täiskasvanutel ja lastel [18]. Kliinikumis osutab LORi erialal tervishoiuteenuseid Kõrvakliinik.

Töös kasutatav LORi statsionaarselt ravitud haigete aegrida katab perioodi jaanuarist 2008 detsembrini 2018 (vt joonis 2.3.1). Kuu keskmine haiglapatsientide arv LORis on aastate võrdluses märkimisväärselt langenud: aastal 2008 raviti kuus keskmiselt 529 haiget, aastal 2017 peaaegu poole vähem, 287 patsienti. 2018. aastal langes see veelgi, 277 patsiendile. Kogu vaadeldava perioodi jooksul oli haigeid vähim 2016. aasta juunis (55) ning enim 2008. aasta oktoobris (635).



Joonis 2.3.1: Otorinolarüngoloogia haigete arv kuus 2008–2018

Ravitud haigete arv kõigub aasta siseselt tugevalt. Suurim absoluutne aastasisene kõikumine oli aastal 2013, kui oktoobris raviti tervelt 488 haiget, detsembris kõigest 108. Paistab silma muster, et suvekuudel (juuni-august) on ravitud haigeid vähem, sügisest kevadeni rohkem, kusjuures kõrghetked on pigem sügisel ja kevadel. Sesoonne sõltuvus perioodiga 12 on näha ka treeningandmestiku autokorrelatsioonidest: 12., 24. (ka 36.) autokorrelatsioon on selgelt nullist erinevad (vt joonis 2.3.2).

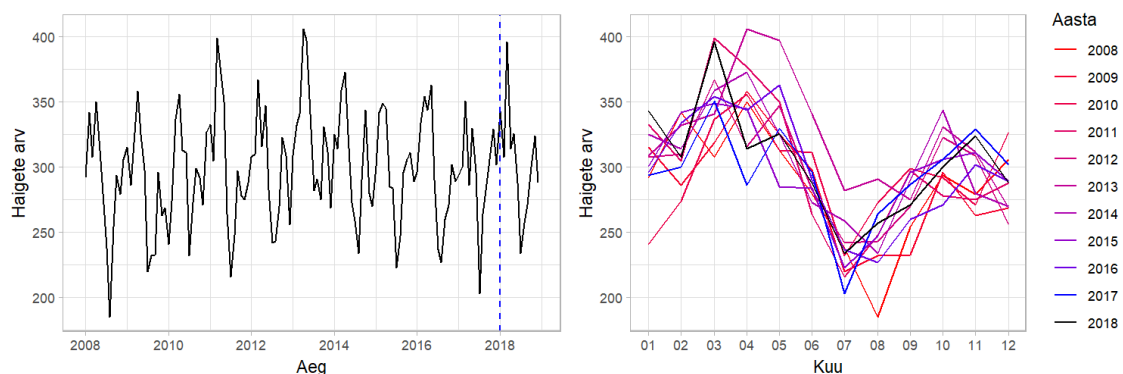


Joonis 2.3.2: LORi treeningandmestik: aegrida, ACF, PACF

Eriala arengukava [18] põhjal on ootuspärane LORi haiglapatsientide arvu üldine langus ja selle arvelt päevaravi osakaalu tõus. Samas on alust arvata, et kliinikumis kui piirkondlikus haiglas peaks mõlemad trendid tulevikus aeglustuma. Seda seetõttu, et ehkki soositakse lihtsamate diagnostiliste ja raviprotseduuride teostamist esmatasandil (perearstid, kohalikud haiglad), üld-, kesk- ja erihaiglates ning päevaravis, jäävad keerukamad ravijuhud ikka piirkondliku haigla statsionaarsesse ravisse.

2.4 Pediaatria andmete kirjeldus

Pediaatria eriala tegeleb lastehaiguste diagnoosimise, ravi, arendus- ja taastusravi, tervise edendamise ja haiguste ennetamisega patsientidel vanuses kuni 18 aastat [19]. Pediaatria erialal teostab tervishoiuteenuseid kliinikumis eeskätt Lastekliinik, aga laste ravimisega tegeletakse ka teistes struktuuriüksustes. Pediaatria katuseriala alla arvatakse antud töös ainult Lastekliiniku tegevus, mis tähendab, et välja jääb laste ravi nt lastekirurgias, onkoloogias, hematoloogias, oftalmoloogias, LORis ja intensiivravis.

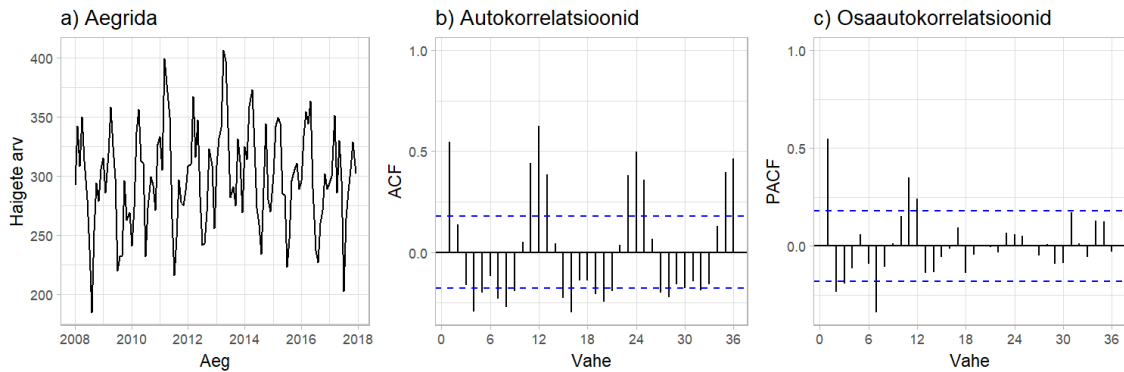


Joonis 2.4.1: Pediaatria haigete arv kuus 2008–2018

Pediaatrias statsionaarselt ravitud haigete aegrida hõlmab perioodi jaanuarist 2008 detsembrini 2018 (vt joonis 2.4.1). Keskmise haigete arv kuus on aastate võrdluses jäänud enam-vähem samaks (keskmiselt 285–324 haiget). Selget trendi ei ole märgata, ehkki perioodi esimeses pooles, s.o aastatel 2008–2013 keskmine ravitud haigete arv mõnevõrra tõusis ning seejärel langes ligikaudu esimese poole keskmisele tasemele. Vähim oli ravitud haigeid 2008. aasta augustis (185), enim 2013. aasta aprillis (406).

Haigete arv kõigub aasta sees märkimisväärselt ning näha on küllalt selge sesoonne muster (perioodiga 12), mis peegeldub ka treeningandmestiku autokorrelatsioonides (vt joonis 2.4.2, lk 22). Seejuures kevadkuudel (märts-mai) on haigeid kõige rohkem, suvekuudel (juuni–august) kõige vähem (sellist kõikumist

saab selgitada näiteks väikelaste infektsioonhaiguste, allergia jmt haiguste sesoonsusega). Suurim aastasisene kõikumine oli aastal 2011, kui juulis raviti 216, märtsis 399 patsienti.



Joonis 2.4.2: Pediaatria treeningandmestik: aegrida, ACF, PACF

Pediaatria eriala arengukava [19] alusel on haigete arvu suhteline muutumatuks jäämine (vähemalt aastatel 2015–2020) ootuspärane. Arengukava ei näe ette ka seda, et lisaks demograafilistele muutustele peaks miski muu oluliselt patsientide arvu lähitulevikus mõjutama.

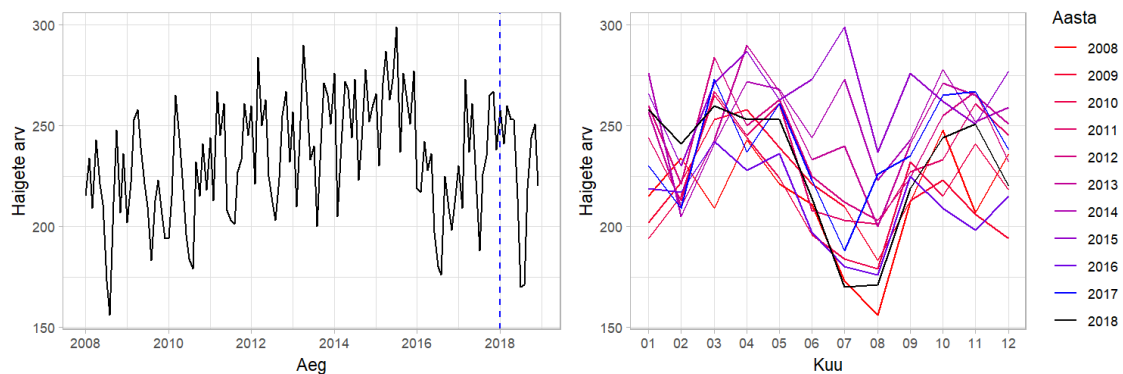
2.5 Sisehaiguste andmete kirjeldus

Sisemeditiini alaerialadeks jaotumine ja sisearsti tööülesanded on riigi- ja haiglaliigispetsiifilised ja suuresti kokkuleppelised ning sõltuvad vastavate spetsialistide olemasolust konkreetses haiglas [20]. Antud töö kontekstis loetakse sisehaiguste katuseriala alla osa kliinikumi Sisekliinikus praktiseeritavatest erialadest: kaasa on arvatud endokrinoloogia, gastroenteroloogia, reumatoloogia ja (üld)sisehaiguste erialad; välja on arvatud infektsioonhaigused ning nefroloogia. Gastroenteroloogia hõlmab seedeelundihaiguste ennetamist, diagnostikat ja ravi; endokrinoloogia tegeleb hormonaalsete häiretega; reumatoloogia tegeleb sidekoe ja tugielundkonna ning ümbritsevate pehmete kudede haigustega; (üld)sisehaiguste eriala hõlmab (muude) erinevate sisehaiguste diagnostikat, ravi ja ennetamist [20], [21], [22], [23].

Sisehaiguste haigete arvus kajastuvad valdavalt täiskasvanud, kuna laste ravimisega nendel erialadel tegeleb Lastekliinik ja need on kaasatud pediaatria eriala alla (vt alapunkt 2.4).

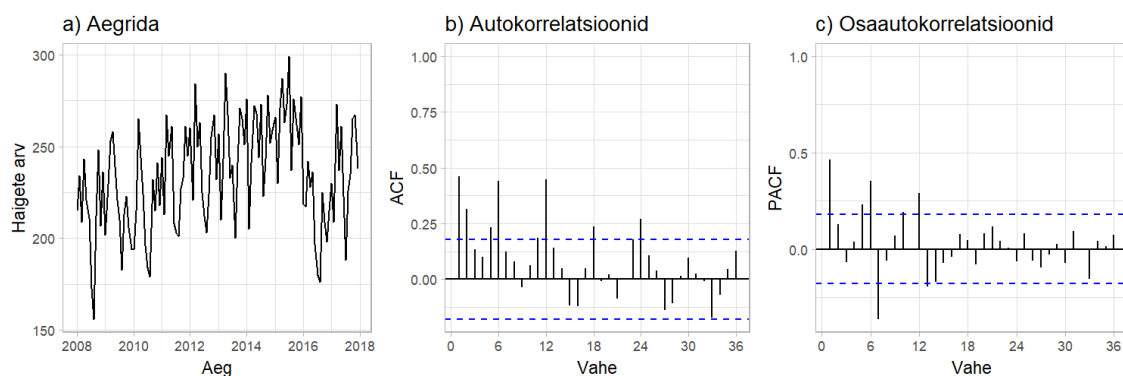
Sisehaiguste haiglapatsientide arvu aegrida on jaanuarist 2008 detsembrini 2018 (vt joonis 2.5.1, lk 23). Kuu keskmine ravitud haigete arv tõusis 214 haigelt 2008. aastal 266 haigeni aastal 2015, see tähendab, et aastas kokku ravitud haigete arv tõusis 2565 patsiendilt 3192 patsiendini. Selle perioodi järjepidevale kasvule

järgnes järsk langus: 2016. aastal raviti kokku 2542 haiget ehk kuus keskmiselt 212 haiget. Sellist muutust haiglapatsientide arvu käitumises selgitab arvatavasti kliiniku kolimine ja struktuurimuudatused, sh päevaravi avamine Sisekliinikus 2016. aastal [24]. Tuleb tähele panna, et muutusejärgne periood on kõigest 2 aastat, mistõttu võib olla ainult nendele andmetele tuginedes raske hinnata statsionaari haigete arvu edasist käitumist. Samas on vähemalt umbes 2016. aasta keskpaigast märgata haigete arvu kasvu mõningast jätkumist.



Joonis 2.5.1: Sisehaiguste haigete arv kuus 2008–2018

Vaadeldavas perioodis on märgata mõningast sesoonset käitumist: valdava osa aastate korral on haigeid vähem suvel, rohkem sügisel ja kevadel (vt joonis 2.5.1). Samas on ka aastaid, mis sellele muustrile ei vasta, näiteks 2015. aasta juulis raviti 299 patsienti, mis on nii selle aasta kui kogu perioodi kõige kõrgem ravitud haigete arv kuus. Valdavalt on siiski juuli ja august aastate madalaima haiglapatsientide arvuga kuud ning perioodi väikseima ravitud haigetega kuu on 2008. aasta august, kus raviti 156 patsienti.



Joonis 2.5.2: Sisehaiguste treeningandmestik: aegrida, ACF, PACF

Mõningase sesoonsuse olemasolule viitavad ka treeningandmestiku autokorrelatsioonid (vt joonis 2.5.2). Nende põhjal võiks öelda, et erinevalt teistest seni vaadeldud ridadest on sisehaiguste ravitud haigete arvu sesoonsuse perioodiks

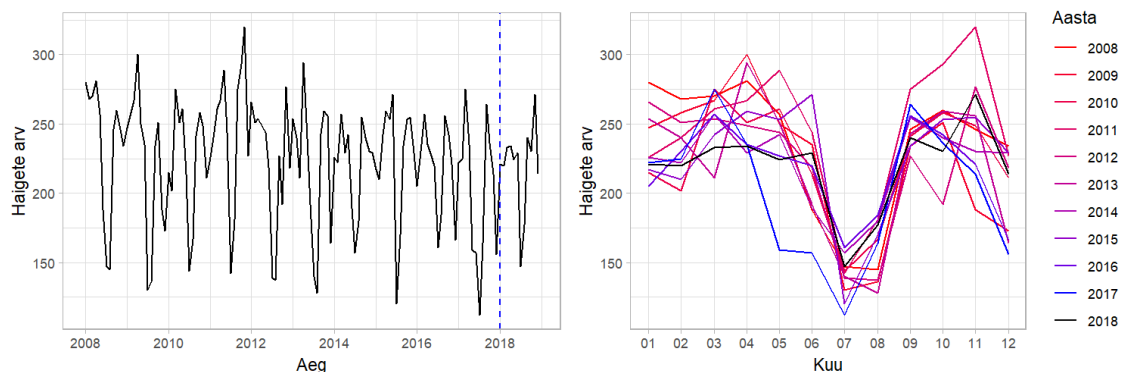
6, ent sesoonsus ei ole väga tugev: autokorrelatsioonid 6 ja 12 on suuremad, aga mitte ühelähedased.

Kõigil selles töös sisehaiguste alla arvatud erialadel on nähtud eriala arengukavade [20], [21], [22], [23] alusel ette haigete arvu jäämist samale tasemele või kasvu vaadeldud perioodis ja lähitulevikus (seda nii statsionaarses kui päevaravis, mis tähendab, et päevaravi avamine kliinikumis pigem ei tähenda haiglapatsientide arvu täielikku stabiliseerumist). Kokku on neil erialadel pigem oodata haigete arvu kasvu.

2.6 Üld- ja plastikakirurgia andmete kirjeldus

Antud töö kontekstis loetakse üld- ja plastikakirurgia (edaspidi ÜPK) katuseriala alla üldkirurgia ja plastikakirurgia eriala, millega tegelevad kliinikumis teatud Kirurgiakliiniku allüksused. Haigete arvus kajastuvad valdavalt täiskasvanud, kuna laste üldkirurgiline ravi kuulub lastekirurgia eriala alla, millega tegeletakse teistes allüksustes.

Plastikakirurgia tegeleb keha nähtavate osade deformatsioonide ja vaegmoodustiste rekonstruktsiooni, reparaatsiooni ja korrigeerimisega [25]. Üldkirurgia hõlmab naha ja pehmete kudede, rinna, seedetrakti, kõhuseina, veresoontesüsteemi ja hormoonsüsteemi haiguste diagnostikat ja kirurgilist ravi (ka siirdamisi) [26]. Vastavalt arengukavale [26] on üldkirurgia alla kaasatud ka kirurgiline onkoloogia. Ehkki Kirurgiakliiniku üldkirurgiaga tegelevates allüksustes teostatakse kasvajate kirurgilist ravi, ei kata see kogu potentsiaalselt üldkirurgia eriala alla loetavat kirurgilist vähiravi, kuna sellega tegeletakse suures mahus ka Hematoloogia-onkoloogiakliinikus.

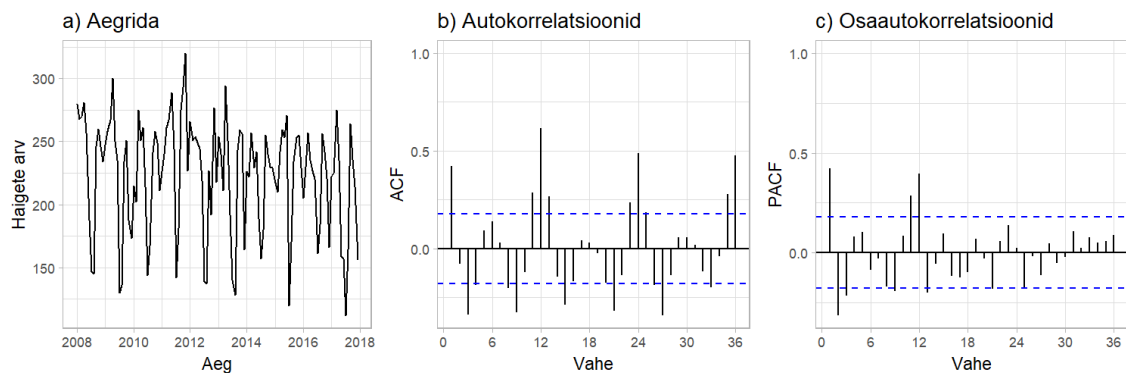


Joonis 2.6.1: Üld- ja plastikakirurgia haigete arv kuus 2008–2018

ÜPK haiglapatsientide arvu aegrida hõlmab perioodi jaanuarist 2008 kuni detsembrini 2018 (vt joonis 2.6.1). Aastas kokku ja seega ka kuus keskmiselt ravitud haigete arvus esineb ehk mõningane langev trend, ent olulist muutust

tasemes siiski näha ei ole. Kuu keskmine ravitud haigete arv on jäänud 202 ja 247 vahele, 2018. aastal oli see 220. Kõige vähem haigeid raviti 2017. aasta juulis (112), kõige rohkem 2011. aasta novembris (320).

Joonistub välja küllaltki tugev sesoonne muster, kus haigete arv langeb juulis-augustis teistest kuudest oluliselt madalamale ning enim haigeid ravitakse kevad- ja sügiskuudel. Sesoonsus perioodiga 12 peegeldub ka treeningandmestiku autokorrelatsioonides, mis on nullist erinevad vahede 12, 24 ja 36 korral (vt joonis 2.6.2).



Joonis 2.6.2: ÜPK treeningandmestik: aegrida, ACF, PACF

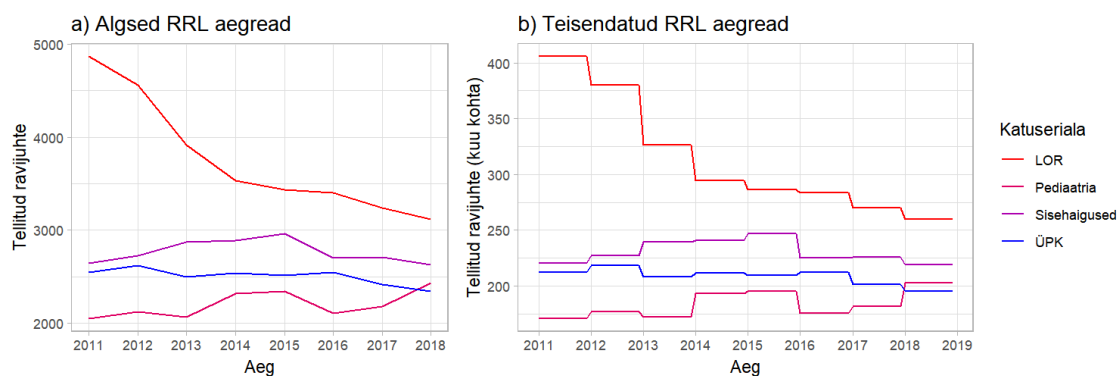
Arengukavadele [25], [26] tuginedes saab öelda, et ka üld- ja plastikakirurgias minnakse erinevate haiguste ravis võimalusel üle päevaravile. Samas jätkub kliinikumis piirkonna patsientide keerulisem ravi (mis toimub haiglaravi korras) ning seetõttu ei pruugi ÜPK haiglapatsientide arvu päevaravi üldine mahu suurendamine oluliselt mõjutada. Kättesaadava aegrea põhjal sellist mõju ka eriti näha ei ole.

2.7 Ravi rahastamise lepingu ravijuhtude andmestike kirjeldus

Ühemõõtmeliste mudelite hinnangute täpsustamiseks püüti töös kaasata mudelitesse RRLi ravijuhtude arvu, mille kohta päriti andmed kliinikumi raviarvelduse osakonnalt (need andmed on kättesaadavad vähemalt jooksva aasta kohta ka haigekassa kodulehel, vt [27]). Töös on kasutatud RRLi ravijuhtude arvu aegridu 2011–2018 aasta alguste seisuga ehk n-ö planeeritavaid või tellitavaid ravijuhtude arve vastaval erialal^{2,3}.

²Aastate 2011–2016 ravijuhtude arv on korrigeeritud hooldaja arvete suhtes, et andmed oleks võrreldavad (puudutab LORi ja pediaatria aegrida).

³RRLis kajastuvate erialade ravijuhtude arvude grupeerimine töös vaadeldavate katuserialade alla on tehtud sarnaselt haigete arvude aegridadele, vt punktid 2.2–2.6 ning lisa 1, tabel L1.1 (lk 57). Kuna teatud tegevused (nt organsiirdamised jmt) ei kajastu RRLis vastava arstliku eriala



Joonis 2.7.1: RRL ravijuhtude arvud aasta alguse seisuga

Nii algsed kui kuu tasandile teisendatud ravijuhtude aegread on toodud joonisel 2.7.1 (kuu tasandile teisendatud aegrea väärtus vastab selleks aastaks tellitud ravijuhtude arvule, mis on jagatud 12-ga).

Jooniselt on näha, et LORi erialal on tellitud ravijuhtude arv kogu vaadeldavas perioodis aasta-aastalt vähenenud: aastal 2011 oli tellitud ravijuhtude arv LORis 4870, aastaks 2018 langes see 3115 ravijuhuni, vähenedes kokku 36% võrra. Teiste erialade puhul nii suurt ühesuunalist muutust ei esine, ehkki ka ÜPK erialal võib täheldada mõningat langevat trendi: 2018. aastal telliti kliinikumilt 2345 ÜPK statsionaarset ravijuhtu, mis on ca 8% võrra vähem kui aastal 2011.

Pediaatria erialal vahelduvad tõusud langustega, ehkki perioodi alguse ja lõpu võrdluses on märgata positiivset muutust: tellitud ravijuhtude arv on tõusnud ligikaudu 19% võrra, s.o 2049lt 2429le. Sisehaiguste erialal on näha mõningat tõusu 2011–2015, kuid aastaks 2018 saavutab tellitud ravijuhtude arv sisehaiguste erialal taas 2011. aasta taseme.

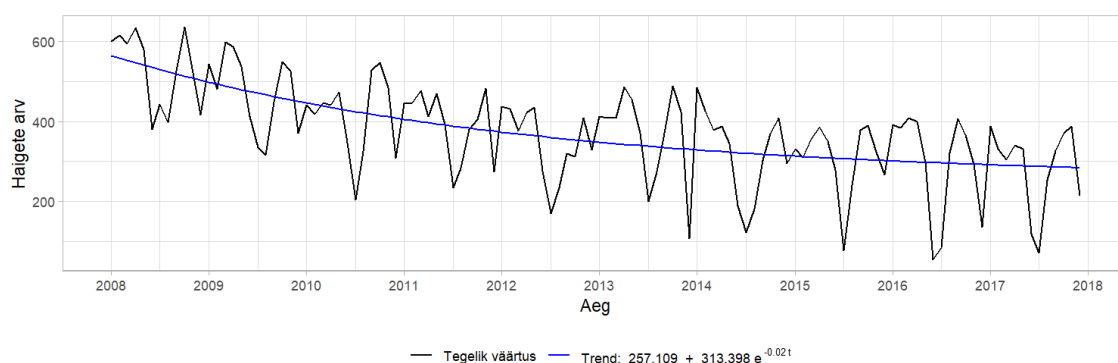
real ning ravijuhtude arvus ei kajastu nt kindlustamata isikute haigusjuhud, siis ei ole ravijuhtude arvud ning haigete arvud üksüheselt võrreldavad.

3 Mudelite hindamine ja sobivuse analüüs

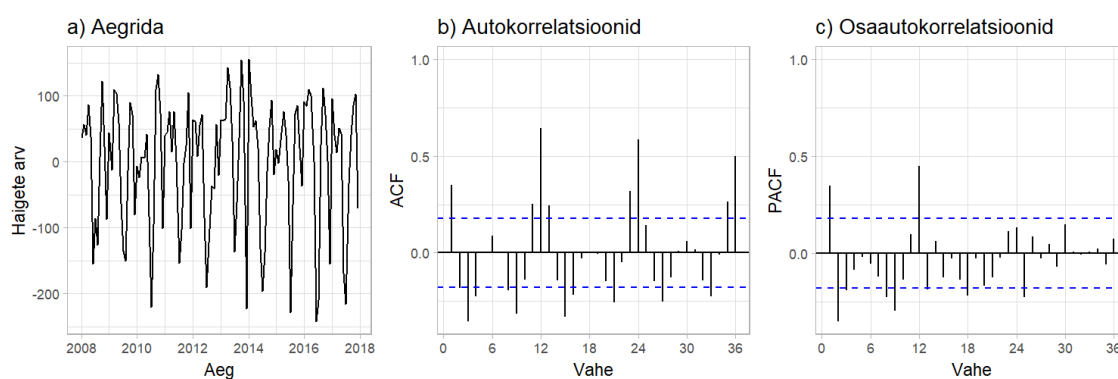
3.1 Otorinolarüngoloogia mudelid

Ühemõõtmelised mudelid

Punktis 2.3 on kirjeldatud LORi aegrea käitumist. Joonisel 2.3.2 (lk 20) kujutatud treeningandmete põhjal on aegreas näha taseme langust, mis tähendab, et aegrida ei saa pidada statsionaarseks. Taustainfole tuginedes on alust eeldada, et tegu on deterministliku stabiliseeruva trendiga. Sellest lähtuvalt eemaldati enne mudelite hindamist esmalt aegreast eksponentsiaalne trend, sobitades reale funktsiooni kujul $f(t) = c_1 + c_2 e^{c_3 t}$. Sobitatud trend on toodud joonisel 3.1.1 ning ülevaade trendi eemaldamisel saadud jääkide aegreast joonisel 3.1.2.



Joonis 3.1.1: LOR: eksponentsiaalne trend



Joonis 3.1.2: LOR: eksponentsiaalse trendi jääkrida

Jääkaegreas esineb sesoonsus perioodiga 12 (vt joonis 3.1.2), mida prooviti arvesse võtta AR(1) ja AR(2) mudeliga sesoonses osas, kuna osautokorrelatsioon 12 on selgelt nullist erinev ning ka osautokorrelatsioon 24 on veidi üle piiri. Et autokorrelatsioon 36 on tugevalt nullist erinev, siis liikuva keskmise mudel sesoonses osas oleks vajanud enam parameetreid.

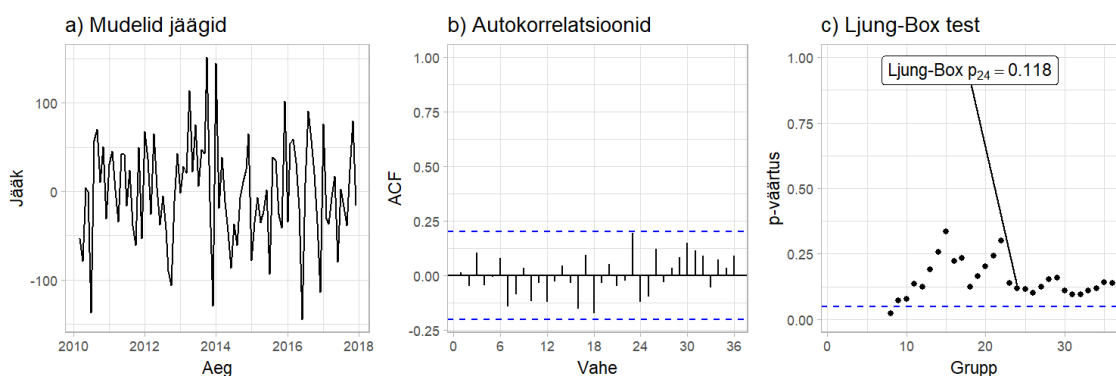
Mittesesoonses osas prooviti AR(2) mudelit, kuna esimesed kaks osaaautokorrelatsiooni on nullist erinevad. Liikuv keskmine mudel mittesesoonses osas vajaks vähemalt 3 või 4 parameetri kaasamist.

Tabel 3.1.1: LORi võimalikud ühemõõtmelised mudelid

Jrk	Mudeli tähis	Mudel	$p_{LB(24)}$
1.	LOR.m1.1	ARIMA(2, 0, 0)(1, 0, 0) ₁₂ eksp. trendi eeldusel	0,004
2.	LOR.m1.2	ARIMA(2, 0, 0)(2, 0, 0) ₁₂ eksp. trendi eeldusel	0,118 *

* – mudel sobib ($p_{LB(24)} > 0,05$)

Kirjeldatud mudelid on ühe parameetri suhtes mittelineaarsed, ent tarkvaras ei ole selliseid mudeleid lihtsasti võimalik sobitada. Seetõttu vaadeldi antud juhul lihtsustatud lähenemist, kus juba hinnatud trendi mudelist fikseeriti parameetri c_3 väärtus ning hinnati uuesti ülejäänud parameetrid (sh c_1 ja c_2). Sobitatud mudelid hindamise järjekorras on toodud tabelis 3.1.1.



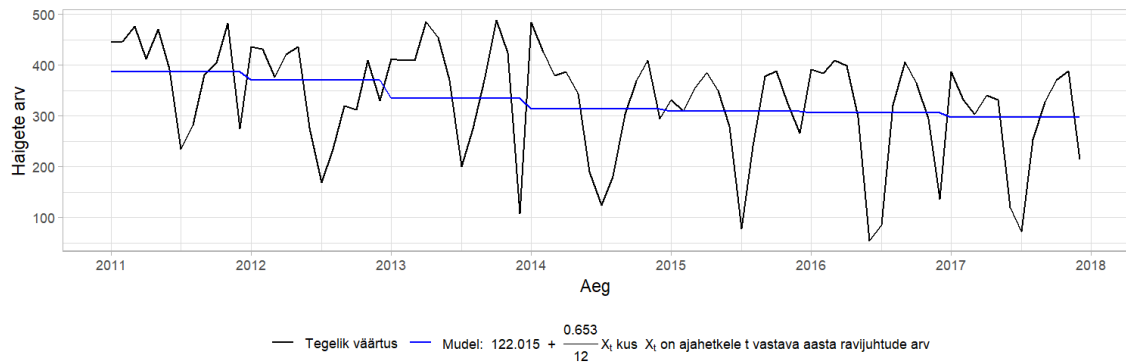
Joonis 3.1.3: Mudeli LOR.m1.2 jääkide diagnostika (ARIMA(2, 0, 0)(2, 0, 0)₁₂ eksponentsiaalse trendi eeldusel)

Grupi 24 jaoks teostatud Ljung-Box testi põhjal ei osutunud mudel LOR.m1.1 sobivaks (vt ka joonis L2.1.1, lk 58). Küll aga sobis andmetega mudel LOR.m1.2 ehk ARIMA(2, 0, 0)(2, 0, 0)₁₂ eksponentsiaalse trendi eeldusel. Selle mudeli jääkide aegrida võib pidada statsionaarseks, kuna ei esine selget taseme ega varieeruvuse muutumist ajas (vt joonis 3.1.3). Jääkide aegrea joonise, autokorrelatsioonide ning Ljung-Box testi põhjal ei ole alust arvata, et jääkide keskvärtus või autokorrelatsioonid erineks nullist.

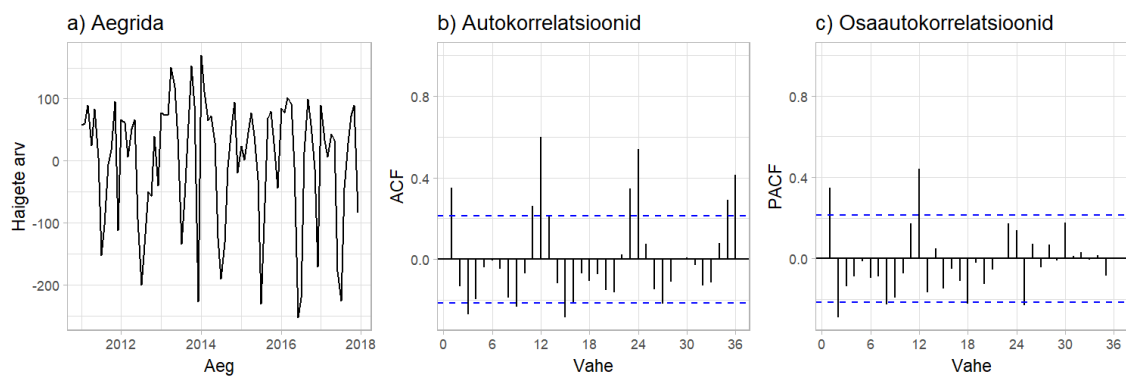
Mitmemõõtmelised mudelid

Kasutades regressorreana RRLile vastavat tellitud ravijuhtude arvu LORi erialal, hinnati esmalt lineaarne regressioonimudel. Selle mudeli sobitatud väärtused on toodud joonisel 3.1.4 (lk 29) ning ülevaade jääkreast joonisel 3.1.5 (lk 29).

Jääkaegrea autokorrelatsioonid ja osautokorrelatsioonid käituvad sarnaselt LORi aegreale, kust on eemaldatud deterministlik trend (vrd joonisega 3.1.2, lk 27). Võimalike mudelite valik jääkidele on seetõttu analoogne trendi eemaldamisel saadud jääkide juhuga: AR(2) mittesesoonses osas, AR(1) või AR(2) sesoonses osas, kus sesoonsuse periood on 12.



Joonis 3.1.4: LOR: lineaarne regressioonimudel



Joonis 3.1.5: LOR: lineaarse regressioonimudeli jääkrida

Sobitatud mudelid hindamise järjekorras on antud tabelis 3.1.2. Analoogselt trendi eemaldamise juhuga, ei osutunud ühe sesoonse autoregressiivse parameetriiga mudel LOR.mx1.1 sobivaks (vt ka joonis L2.1.2, lk 58).

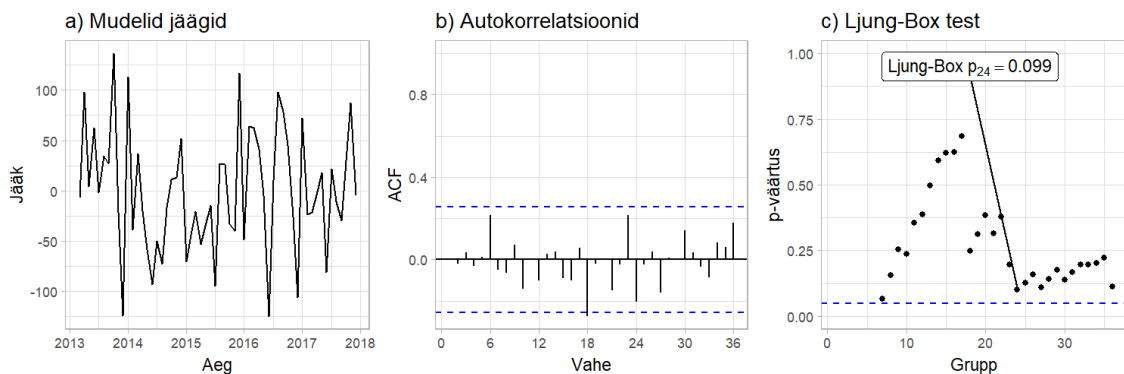
Tabel 3.1.2: LORi võimalikud mitmemõõtmelised mudelid

Jrk	Mudeli tähis	Mudel	$p_{LB(24)}$
1.	LOR.mx1.1	lineaarne regressioonimudel	0,031
		ARIMA(2, 0, 0)(1, 0, 0) ₁₂ vigadega	
2.	LOR.mx1.2	lineaarne regressioonimudel	0,099 *
		ARIMA(2, 0, 0)(2, 0, 0) ₁₂ vigadega	

* – mudel sobib ($p_{LB(24)} > 0,05$)

Mudeli LOR.mx1.2 jääkide aegreas (joonis 3.1.6, lk 30) paistavad silma järsud tõusud iga aasta alguses, mis tuleneb tõenäoliselt regressorrea tükiti konstantsest

kujust. Samas jääkide autokorrelatsioonide põhjal ei paista silma kunstlikult tekkinud täiendavat aastase perioodiga sesoonsust. Jääkide keskväärts paistab olevat null, ehkki esineb perioode, mille vältel jäägid jäävad nullist pikemalt eemale (nt 2014. aastal ca 6 kuud). Jääkide autokorrelatsioon vahe 18 juures on piiri lähendal, ent üldiselt paistavad autokorrelatsioonid olevat võrdsed nulliga, mida kinnitavad ka Ljung-Box testi suured olulisuse tõenäosused. Mõningase reservatsiooniga (kasutatud regressorrea iseloomu tõttu) võib mudeli LOR.mx1.2 ehk lineaarse regressioonimudeli $ARIMA(2, 0, 0)(2, 0, 0)_{12}$ vigadega lugeda andmetega sobivaks.



Joonis 3.1.6: Mudeli LOR.mx1.2 jääkide diagnostika (lineaarne regressioonimudel $ARIMA(2, 0, 0)(2, 0, 0)_{12}$ vigadega)

3.2 Pediaatria mudelid

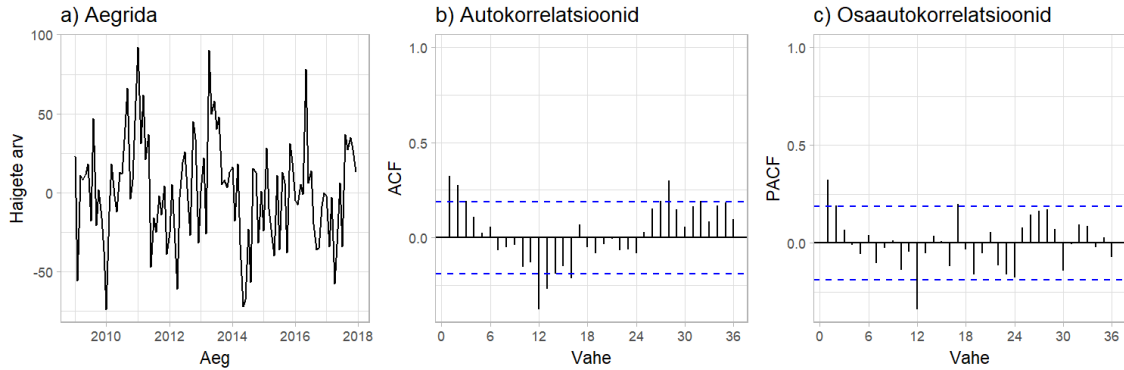
Ühemõõtmelised mudelid

Joonise 2.4.2 (lk 22) põhjal on pediaatria aegreas pigem tugev sesoonsus (perioodiga 12). Kuna aegrida ei ole väga pikk, siis ainult andmete põhjal pole selge, kas sesoonsus tuleks arvesse võtta sesoonse liikme kaasamisega või peaks rida sesoonselt diferentsima. Seetõttu uuritakse mõlemat juhtu.

Kui vaadelda diferentsimata andmeid (joonis 2.4.2, lk 22), siis võiks osaaautokorrelatsioonide põhjal sesoonses osas kasutada $AR(1)$ mudelit. Samas osaaautokorrelatsioon vahe 7 juures on küllalt tugevalt üle piiri, mida võiks vajadusel proovida arvesse võtta mudeliga, kus sesoonsuse periood on 6 ning kaasatud on 2 sesoonset autoregressiivset parameetrit. Mittesesoonses osas võiks sobida autoregressiivne mudel parameetrite arvuga 1 või 2, kuna esimene osaaautokorrelatsioon selgelt nullist erinev, aga ka teine osaaautokorrelatsioon on mõnevõrra üle piiri.

Kuna taustainfole tuginedes võib oletada, et sesoonsus on püsiv, siis prooviti mudelit leida ka sesoonsele muudule. Ülevaade muutude treeningaegreast on

toodud joonisel 3.2.1. Sellest järeldub, et sesoonses osas võiks proovida AR(1) või MA(1) mudelit, kuna 12. autokorrelatsioon ja 12. osaautokorrelatsioon on piiridest väljas. Mittesesoonses osas võiks sobida näiteks AR(1) mudel, kuna 1. osaautokorrelatsioon on selgelt nullist erinev.



Joonis 3.2.1: Sesoonselt diferentsitud ($D = 1, s = 12$) pediatría treeningandmestik

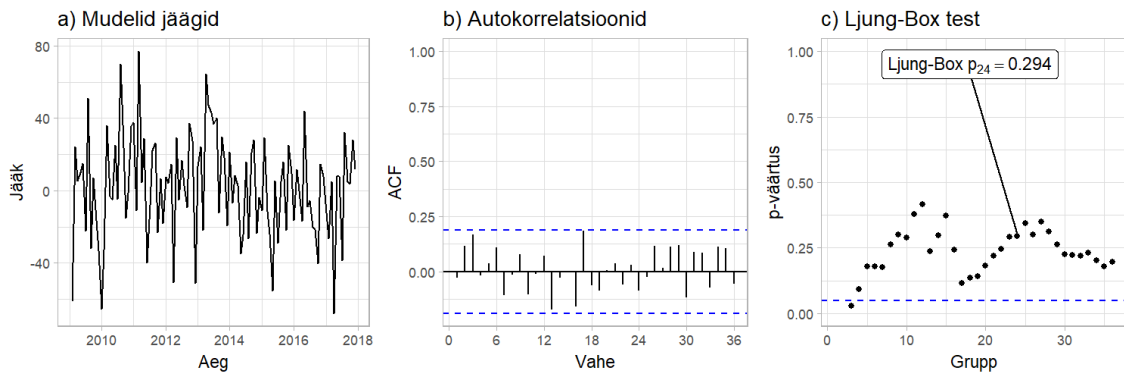
Sobitatud mudelid (sobitamise järjekorras) koos Ljung-Box testi olulisuse tõenäosusega grupile 24 on toodud tabelis 3.2.1. Kõik originaalandmetele valitud mudelid osutusid sobivaks, kusjuures mudelite jääkide käitumine on küllaltki sarnane: jääkide aegrida näib olevat statsionaarne keskväertusega 0, jääkide 16. autokorrelatsioon on piiridest väljas, ent Ljung-Box testi olulisuse tõenäosused on suured grupini 36 (vt lk 59–60, joonised L2.2.1, L2.2.2, L2.2.3, L2.2.4).

Tabel 3.2.1: Pediatría võimalikud ühemõõtmelised mudelid

Jrk	Mudeli tähis	Mudel	$p_{LB(24)}$
1.	PED.m1.1	ARIMA(1, 0, 0)(1, 0, 0) ₁₂	0,263 *
2.	PED.m1.2	ARIMA(1, 0, 0)(2, 0, 0) ₆	0,238 *
3.	PED.m1.3	ARIMA(2, 0, 0)(1, 0, 0) ₁₂	0,351 *
4.	PED.m1.4	ARIMA(2, 0, 0)(2, 0, 0) ₆	0,353 *
1.	PED.m2.1	ARIMA(1, 0, 0)(1, 1, 0) ₁₂	0,006
2.	PED.m2.2	ARIMA(1, 0, 0)(0, 1, 1) ₁₂	0,294 *

* – mudel sobib ($p_{LB(24)} > 0,05$)

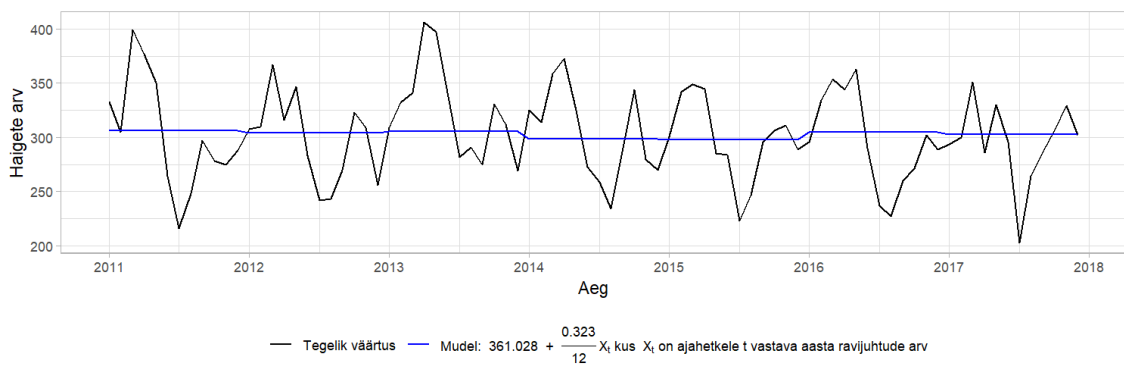
Sesoonse diferentsiga mudelitest ei sobinud andmetega esimesena proovitud mudel PED.m2.1, mille jäägid ei ole Ljung-Box testi põhjal sõltumatud (vt ka joonis L2.2.5, lk 60). Küll aga osutus sobivaks mudel PED.m2.2 ehk ARIMA(1, 0, 0)(0, 1, 1)₁₂, mille jäägid paistavad joonise 3.2.2 (lk 32) põhjal olevat mittekorreleeritud ning keskväertusega 0.



Joonis 3.2.2: Mudeli PED.m2.2 jääkide diagnostika (ARIMA(1, 0, 0)(0, 1, 1)₁₂)

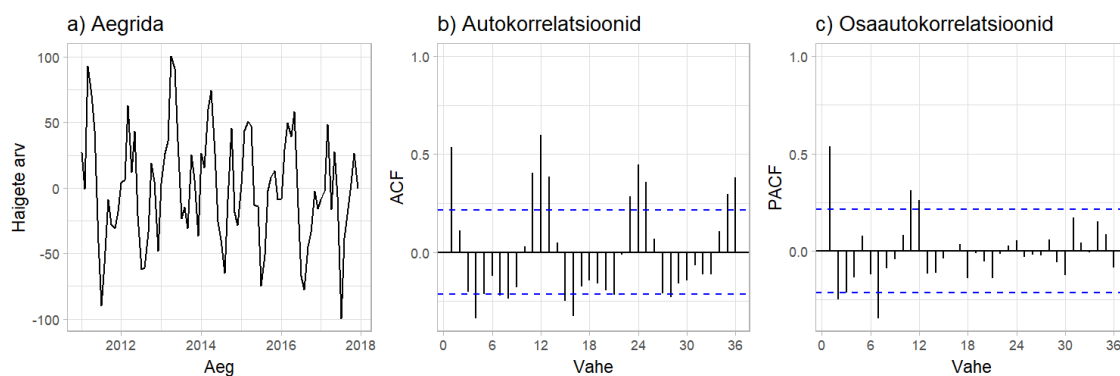
Mitmemõõtmelised mudelid

Kasutades regressorina RRLile vastavat tellitud ravijuhtude arvu aegrida pediatría erialal, hinnati esmalt lineaarne regressioonimudel. Selle mudeli hinnangud on toodud joonisel 3.2.3 ning ülevaade jääkreast joonisel 3.2.4, lk 33.



Joonis 3.2.3: Pediatría: lineaarne regressioonimudel

Jääkaegrea autokorrelatsioonid ja osaaautokorrelatsioonid käituvad sarnaselt algsele pediatría aegreale, mistõttu on jääkide mudelite valimine analoogne ühemõõtmelise juhuga. Hinnatud mudelid (sobitamise järjekorras) on toodud tabelis 3.2.2, lk 33. Kõik valitud mudelid osutusid sobivaks ning jääkide diagnostika on kõigi mudelite korral küllaltki sarnane (vt ka joonis 3.2.5, lk 33; joonised L2.2.6, L2.2.7, L2.2.8, lk 60–61).

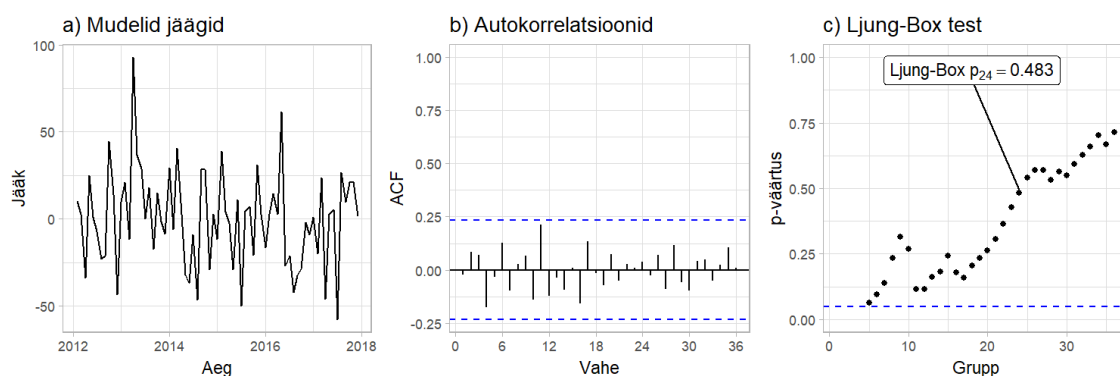


Joonis 3.2.4: Pediaatria: lineaarse regressioonimudeli jääkrida

Tabel 3.2.2: Pediaatria võimalikud mitmemõõtmelised mudelid

Jrk	Mudeli tähis	Mudel	$p_{LB(24)}$
1.	PED.mx1.1	lineaarne regressioonimudel ARIMA(1, 0, 0)(1, 0, 0) ₁₂ vigadega	0,483 *
2.	PED.mx1.2	lineaarne regressioonimudel ARIMA(1, 0, 0)(2, 0, 0) ₆ vigadega	0,422 *
3.	PED.mx1.3	lineaarne regressioonimudel ARIMA(2, 0, 0)(1, 0, 0) ₁₂ vigadega	0,514 *
4.	PED.mx1.4	lineaarne regressioonimudel ARIMA(2, 0, 0)(2, 0, 0) ₆ vigadega	0,474 *

* – mudel sobib ($p_{LB(24)} > 0,05$)



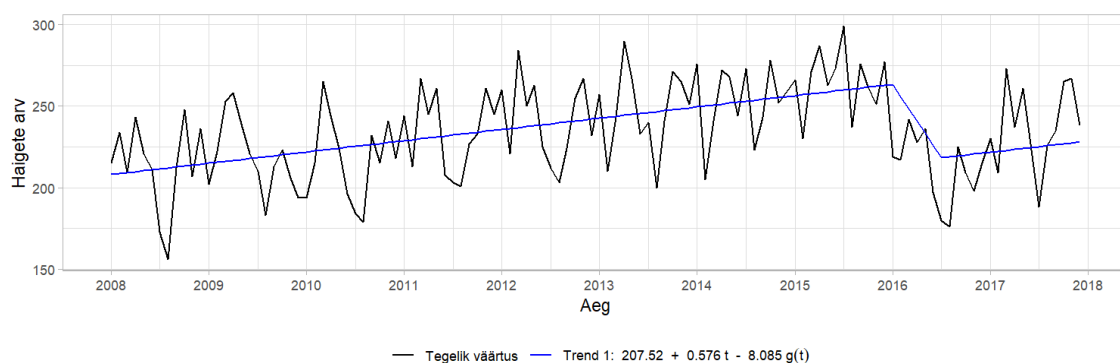
Joonis 3.2.5: Mudeli PED.mx1.1: jääkide diagnostika (lineaarne regressioonimudel ARIMA(1, 0, 0)(1, 0, 0)₁₂ vigadega)

3.3 Sisehaiguste modelid

Ühemõõtmelised modelid

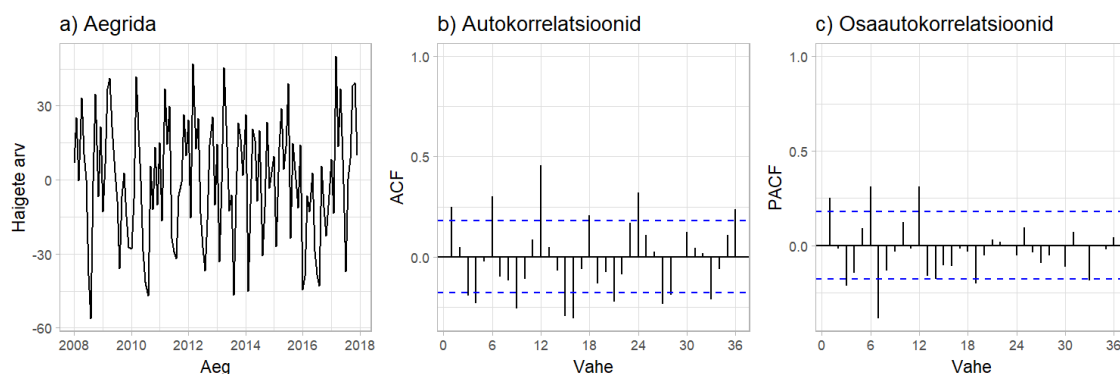
Sisehaiguste aegreas on taustainfo põhjal alust eeldada deterministlikku kasvavat trendi, mis peaks jätkuma ka 2016. aasta esimeses pooles toimuva taseme languse järel sarnaselt (vt joonis 2.5.2, lk 23). Sellest lähtuvalt hinnati esmalt trendi mudel, sobitades reale funktsiooni kujul $f(t) = c_1 + c_2t + c_3g(t)$, kus

$$g(t) = \begin{cases} 0, & t \leq 97 \text{ (jaanuar 2016)} \\ t - 97, & 97 < t < 103 \\ 6, & t \geq 103 \text{ (juuli 2016)} \end{cases}.$$



Joonis 3.3.1: Sisehaigused: trend 1

See tähendab, et eeldati, et trendisirge tõus on kummalgi pool muutuse perioodi ühesugune (c_2), aga muutub tase: kuni jaanuarini 2016 eeldati aegreal üht taset (c_1), perioodis veebruar–juuni 2016 toimus muutus ning uus tase ($c_1 + 6c_3$) saavutati juulis 2016. Sobitatud trend on toodud joonisel 3.3.1.



Joonis 3.3.2: Sisehaigused: trendi 1 jääkrida

Trendi 1 eemaldamisel saadud jääkrea ülevaade on toodud joonisel 3.3.2. Jääkreas esineb sesoonsus perioodiga 6, mida prooviti arvesse võtta AR(2) mudeliga

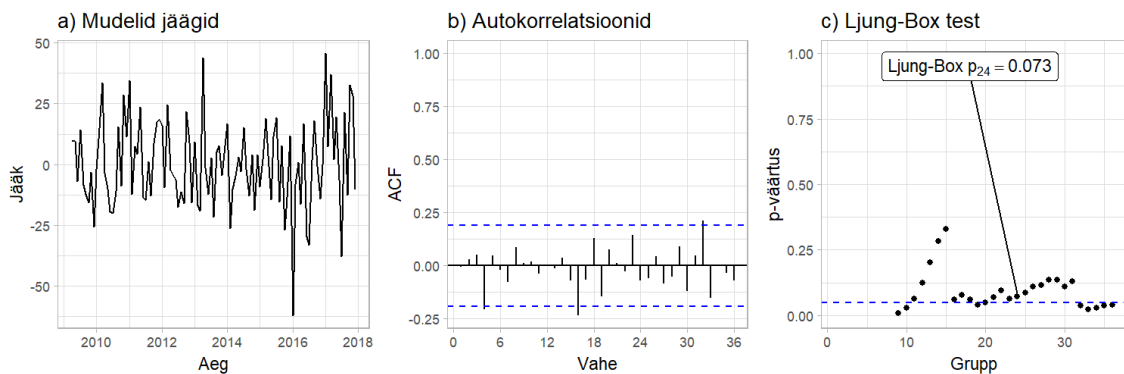
sesoonses osas (kuna osaaautokorrelatsioonid 6 ja 12 erinevad nullist). Liikuva keskmise mudel sesoonses osas vajanuks rohkemate parameetrite hindamist. Mittesesoonses osas otsustati esmalt proovida AR(1) mudelit, kuna kolmas osaaautokorrelatsioon on vaid veidi piiridest väljas. See mudel osustus ebasobivaks, misjärel prooviti AR(3) mudelit.

Tabel 3.3.1: Sisehaiguste võimalikud ühemõõtmelised mudelid

Jrk	Mudeli tähis	Mudel	$p_{LB(24)}$
1.	SISE.m1.1	ARIMA(1, 0, 0)(2, 0, 0) ₆ trendi 1 eeldusel	0,018
2.	SISE.m1.2	ARIMA(3, 0, 0)(2, 0, 0) ₆ trendi 1 eeldusel	0,137 *
1.	SISE.m2.1	ARIMA(1, 0, 0)(2, 0, 0) ₆ trendi 2 eeldusel	0,018
2.	SISE.m2.2	ARIMA(3, 0, 0)(2, 0, 0) ₆ trendi 2 eeldusel	0,028
3.	SISE.m2.3	ARIMA(4, 0, 0)(2, 0, 0) ₆ trendi 2 eeldusel	0,084 *

* – mudel sobib ($p_{LB(24)} > 0,05$)

Mõlemad trendi 1 eeldusel proovitud mudelid, SISE.m1.1 ja SISE.m1.2, koos grupi 24 jaoks teostatud Ljung-Box testi olulisuse tõenäosusega on toodud tabelis 3.3.1. Osutub, et esimene neist mudelitest ei sobi andmetega, kuna leidub mittenullilisi autokorrelatsioone, mida kinnitab Ljung-Box testi väike olulisuse tõenäosus (vt ka joonis L2.3.1, lk 62).



Joonis 3.3.3: Mudeli SISE.m1.2 jääkide diagnostika (ARIMA(3, 0, 0)(2, 0, 0)₆ trendi 1 eeldusel)

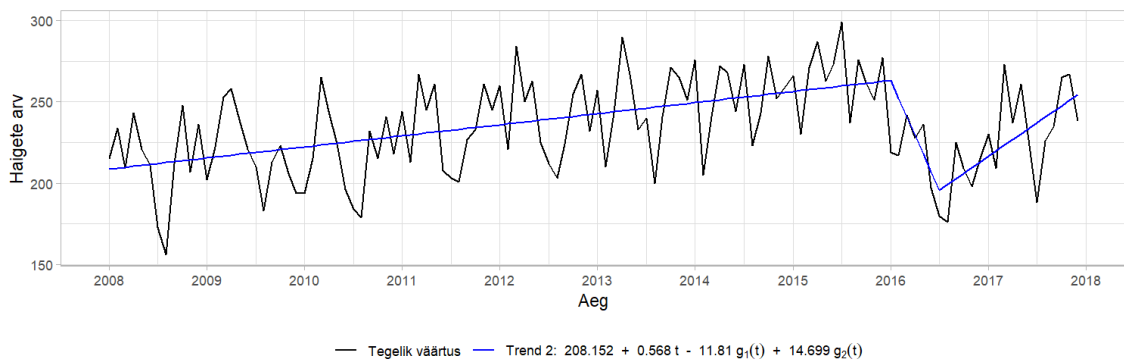
Mudeli SISE.m1.2 jääkide aegreas paistab silma suur negatiivne viga 2016. aasta jaanuaris, mis vastab üleminekuperioodi algusele (vt joonis 3.3.3). Mõned jääkide autokorrelatsioonid on küllaltki piiripealsed või ka piiridest väljas, kuid Ljung-Box test ei lükka (vähemalt olulisuse nivool 0,05) ümber hüpoteesi, et autokorrelatsioonid kuni vaheni 24 (isegi kuni 31) on võrdsed nulliga. Seega võib reservatsiooniga lugeda mudeli SISE.m1.2 ehk mudeli ARIMA(3, 0, 0)(2, 0, 0)₆ trendi 1 eeldusel sobivaks.

Trendi 1 eeldamine sisehaiguste aegreas ei vasta päriselt taustainfole: kuigi

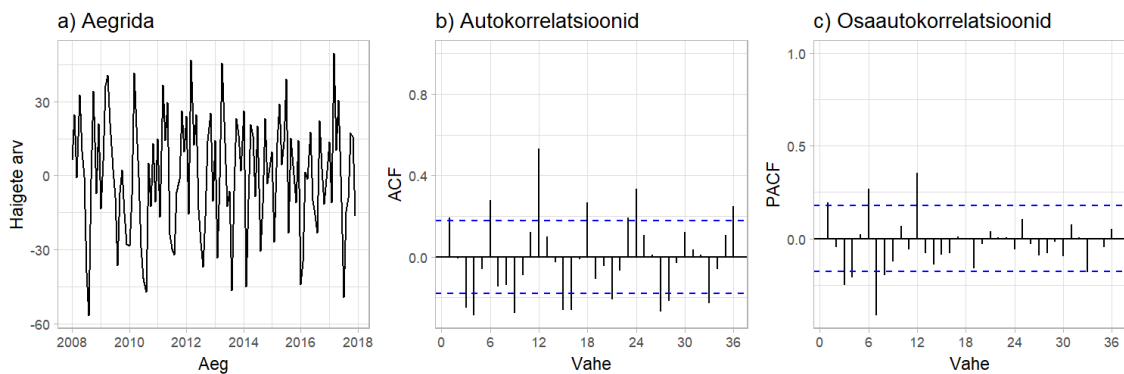
muutusele järgnevas perioodis on oodata üldist haigete arvu kasvu, toimub edaspidi haigete jagunemine haiglaravi ja päevaravi vahel. See tähendab, et tõus ei pruugi olla muutusejärgselt päris samasugune nagu muutusele eelnevas perioodis, vaid võiks pigem olla väiksem. Sellest lähtuvalt prooviti ka lähenemist, kus lubati trendisirge täielikku muutumist enne ja pärast muutust, st sobitati aegreale funktsioon kujul $f(t) = c_1 + c_2t + c_3g_1(t) + c_4g_2(t)$, kus

$$g_1(t) = (t - 97)I_{t > 97} \text{ (jaanuar 2016),}$$

$$g_2(t) = (t - 103)I_{t > 103} \text{ (juuli 2016).}$$



Joonis 3.3.4: Sisehaigused: trend 2

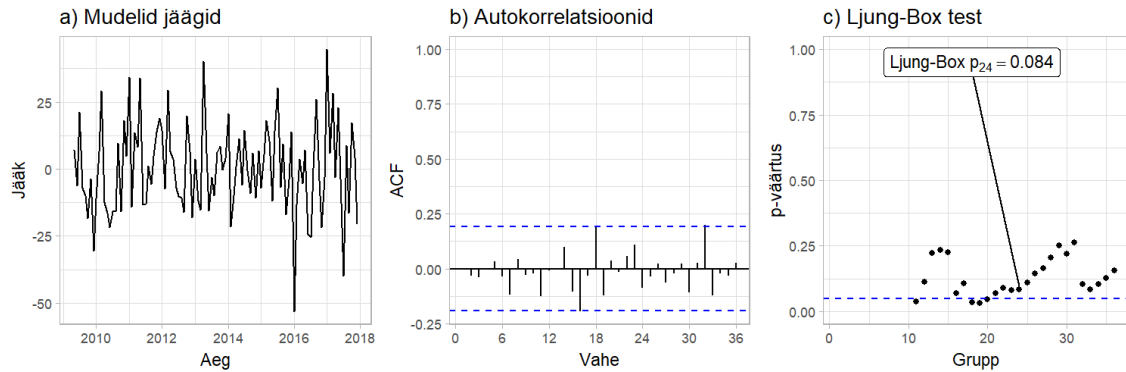


Joonis 3.3.5: Sisehaigused: trendi 2 jääkrida

Sobitatud alternatiivne trend (trend 2) on antud joonisel 3.3.4 ning ülevaade jääkaegreast, mis on saadud selle trendi eemaldamisel, on toodud joonisel 3.3.5. Nagu ka algselt aegrea jooniselt võis arvata, on muutusejärgse perioodi andmeid liiga vähe, et tõusu parameetrit tegelikkusele lähedaselt hinnata: tõus hoopis suureneb muutuse perioodi järgselt. See ei ole kooskõlas taustainfoga, mistõttu võib arvata, et antud trendi eeldusel ei ole tõenäoliselt võimalik head prognoosimodelit leida.

Trendi 2 jääkaegrea autokorrelatsioonid ja osautokorrelatsioonid käituvad

analoogselt trendi 1 juhuga. Esimesed võimalikud mudelid SISE.m2.1 ja SISE.m2.2 on sama tüüpi nagu trendi 1 juhul (vt tabel 3.3.1, lk 35). Antud mudelite jääkide diagnostika on toodud lk 62, joonistel L2.3.2 ja L2.3.3, millele tuginedes tuli lugeda need mudelid ebasobivaks, kuna jääkide autokorrelatsioonid erinevad nullist.



Joonis 3.3.6: Mudeli SISE.m2.3 jääkide diagnostika ($\text{ARIMA}(4, 0, 0)(2, 0, 0)_6$ trendi 2 eeldusel)

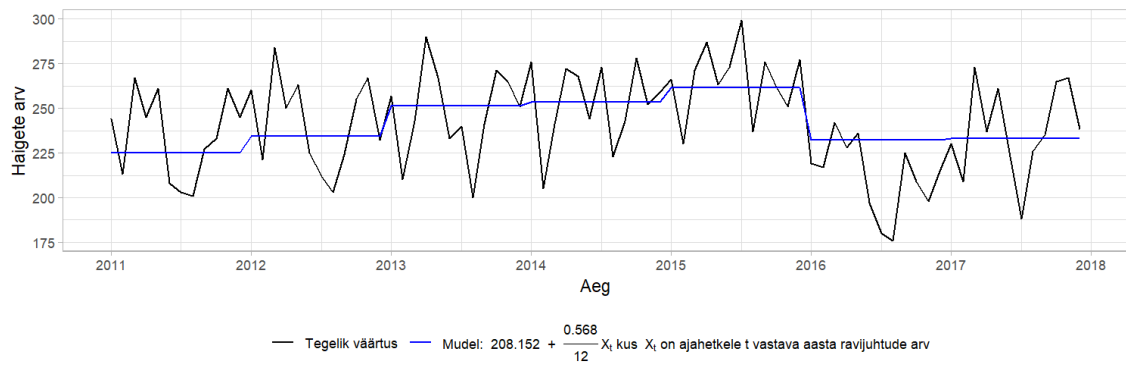
Seejärel prooviti lisada mittesoonsesse osasse täiendav parameeter, saades mudeli SISE.m2.3 ehk mudeli $\text{ARIMA}(4, 0, 0)(2, 0, 0)_6$ trendi 2 eeldusel. Selle võib lugeda andmetega sobivaks (vt joonis 3.3.6). Sarnaselt trendi 1 eeldusel saadud mudelile on ka selle mudeli korral jääk 2016. aasta jaanuaris teiste jääkidega võrreldes küllaltki suur. Mõned autokorrelatsioonid on piiri lähedal, aga need võib lugeda nulliks, millele annavad kinnitust ka Ljung-Box testi pigem suured olulisuse tõenäosused isegi kuni grupini 36. Antud tulemusse tuleks siiski suhtuda reservatsiooniga, arvestades, et trendi 2 eeldusel loodud mudelite näol on tõenäoliselt tegu ülesobitamise.

Mitmemõõtmelised mudelid

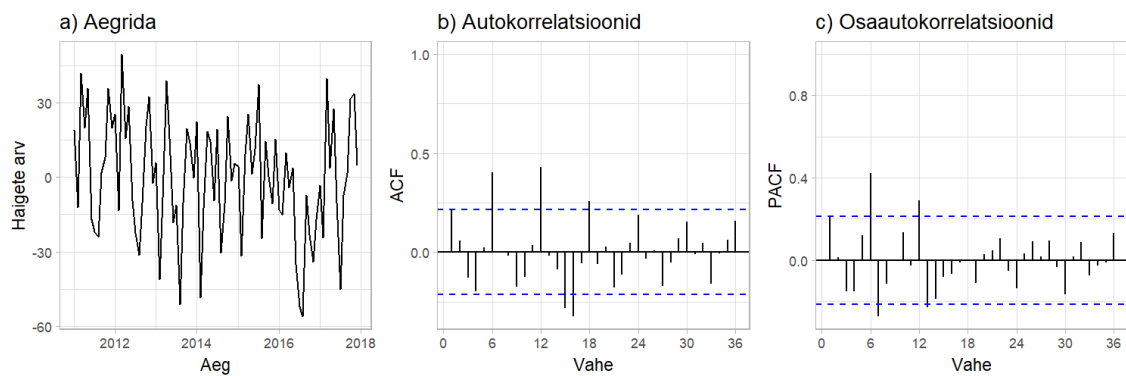
Kasutades regressorina RRLi ravijuhtude arvu aegrida sisehaiguste erialal, hinnati esmalt lineaarne regressioonimudel. Selle mudeli hinnangud on toodud joonisel 3.3.7 (lk 38) ning jääkrea ülevaade joonisel 3.3.8 (lk 38).

Jääkaegreas on sesoonsus perioodiga 6, mida otsustati proovida arvesse võtta $\text{AR}(2)$ mudeliga sesoonses osas (kuna osautokorrelatsioonid 6 ja 12 erinevad nullist). Mittesoonses osas prooviti mudeleid $\text{AR}(1)$ ja $\text{MA}(1)$, kuna nii autokorrelatsioon kui osautokorrelatsioon 1 on veidi üle piiri ja neile vahetult järgnevad saab lugeda võrdseks nulliga.

Sobitatud mudelid hindamise järjekorras on toodud tabelis 3.3.2, lk 38. Mõlemad mudelid osutusid sobivaks (vt ka joonis 3.3.9, lk 38, joonis L2.3.4, lk 63).



Joonis 3.3.7: Sisehaigused: lineaarne regressioonimudel

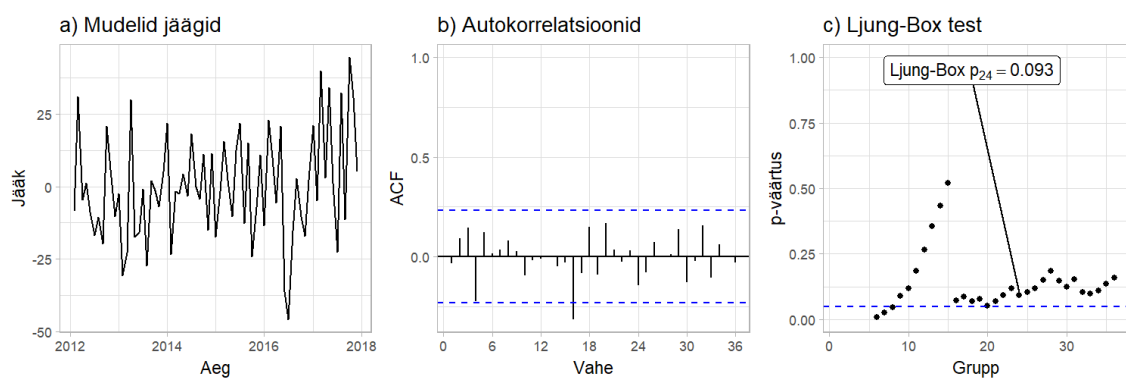


Joonis 3.3.8: Sisehaigused: lineaarse regressioonimudeli jääkrida

Tabel 3.3.2: Sisehaiguste võimalikud mitmemõõtmelised mudelid

Jrk	Mudeli tähis	Mudel	$p_{LB(24)}$
1.	SISE.mx1.1	lineaarne regressioonimudel ARIMA(1, 0, 0)(2, 0, 0) ₆ vigadega	0,093 *
2.	SISE.mx1.2	lineaarne regressioonimudel ARIMA(0, 0, 1)(2, 0, 0) ₆ vigadega	0,084 *

* – mudel sobib ($p_{LB(24)} > 0,05$)



Joonis 3.3.9: Mudeli SISE.mx1.1 jääkide diagnostika (lineaarne regressioonimudel ARIMA(1, 0, 0)(2, 0, 0)₆ vigadega)

3.4 Üld- ja plastikakirurgia mudelid

Ühemõõtmelised mudelid

ÜPK aegreas esineb küllaltki tugev sesoonsus, eriti paistab see silma aastaste miinimumide osas (vt joonis 2.6.2, lk 25). Samas autokorrelatsioonide põhjal otsustades ei ole väga selge, et tuleks rida sesoonselt diferentsida. Seetõttu prooviti esmalt leida mudelit diferentsimata aegreale.

Diferentsimata aegreale mudeli sobitamisel püüti sesoonsus arvesse võtta autoregressiivse sesoonse mudeli abil (kuna liikuva keskmise mudelid vajanuks rohkem parameetreid). Prooviti nii AR(1) kui AR(2) mudelit sesoonses osas (perioodiga 12). Kuna nendega sobivat mudelit ei leitud, siis prooviti sesoonses osas ka segamudelit ARMA(1,1), mis võiks samuti autokorrelatsioonide ja osaautokorrelatsioonide käitumise põhjal sobida. Mittesoonses osas katsetati autoregressiivset mudelit 2 ja 3 parameetriga.

Diferentsimata andmete hinnatud mudelid sobitamise järjekorras koos Ljung-Box testi olulisuse tõenäosustega grupi 24 korral on toodud tabelis 3.4.1 (mudelid ÜPK.m1.1–ÜPK.m1.5). Ainus sobiv mudel oli ÜPK.m1.5 ehk ARIMA(2, 0, 0)(1, 0, 1)₁₂ (vt ka joonis L2.4.1, lk 64).

Tabel 3.4.1: ÜPK võimalikud ühemõõtmelised mudelid

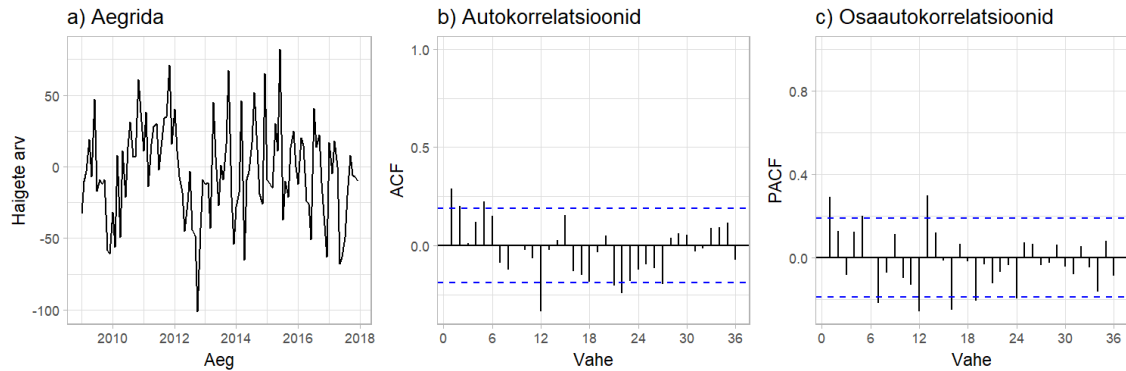
Jrk	Mudeli tähis	Mudel	$p_{LB(24)}$
1.	ÜPK.m1.1	ARIMA(2, 0, 0)(1, 0, 0) ₁₂	0,015
2.	ÜPK.m1.2	ARIMA(3, 0, 0)(1, 0, 0) ₁₂	0,025
3.	ÜPK.m1.3	ARIMA(2, 0, 0)(2, 0, 0) ₁₂	0,043
4.	ÜPK.m1.4	ARIMA(3, 0, 0)(2, 0, 0) ₁₂	0,040
5.	ÜPK.m1.5	ARIMA(2, 0, 0)(1, 0, 1) ₁₂	0,082 *
7.	ÜPK.m2.1	ARIMA(1, 0, 0)(1, 1, 0) ₁₂	0,013
8.	ÜPK.m2.2	ARIMA(0, 0, 1)(1, 1, 0) ₁₂	0,003
9.	ÜPK.m2.3	ARIMA(1, 0, 0)(0, 1, 1) ₁₂	0,156 *
10.	ÜPK.m2.4	ARIMA(0, 0, 1)(0, 1, 1) ₁₂	0,062 *

* – mudel sobib ($p_{LB(24)} > 0,05$)

Järgmisena prooviti otsida mudelit sesoonsele muudule. Ülevaade muutude aegreast on antud joonisel 3.4.1 (lk 40). Esineb pikemaid perioode, mil aegrea väärtused jäävad üldisest tasemest kaugemale (nt aasta 2013) või võib esineda mõningane lokaalne trend (aastatel 2010–2012), kuid ei ole ka selge, et seda rida ei võiks pidada statsionaarseks.

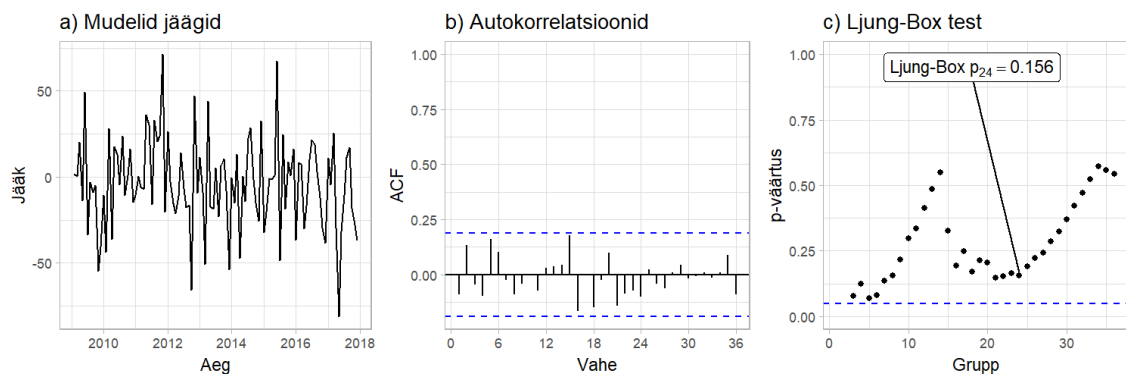
Muutude rea autokorrelatsioonide ja osaautokorrelatsioonide põhjal otsustati

proovida sesoonses osas nii AR(1) kui MA(1) mudelit, kuna ainult 12. autokorrelatsioon ja 12. osautokorrelatsioon on selgelt piiridest väljas. Mittesoonses osas sobitati AR(1) ja MA(1) mudelit. Kõik sesoonse diferentsiga mudelid (ÜPK.m2.1–ÜPK.m2.4) on toodud tabelis 3.4.1, lk 39.



Joonis 3.4.1: Sesoonselt diferentsitud ($D = 1, s = 12$) ÜPK treeningandmestik

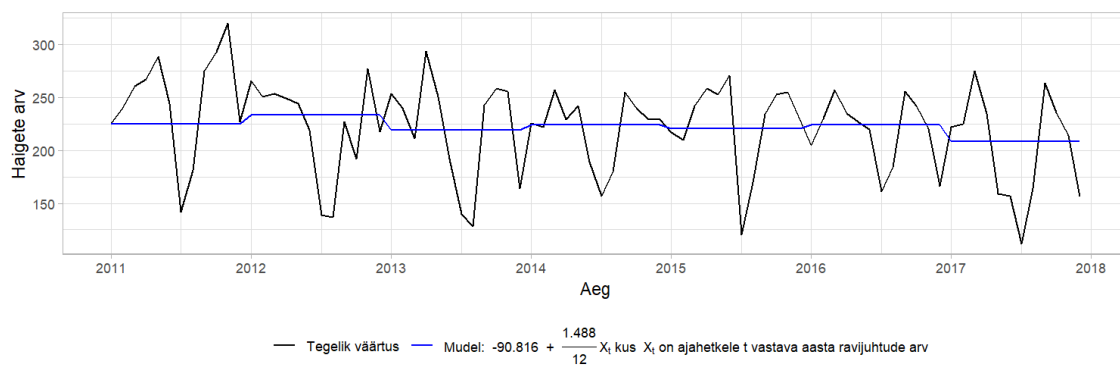
Autoregressiivse sesoonse osaga mudelid ÜPK.m2.1 ja ÜPK.m2.2 osutusid mitesobivaks (vt ka lk 64, joonised L2.4.2, L2.4.3). Liikuva keskmise tüüpi sesoonse osaga mudelid ÜPK.m2.3 ja ÜPK.m2.4 saab lugeda sobivateks. Viimaste mudelite jääkide aegread tunduvad olevat statsionaarsed keskvärtusega 0. Mudel ÜPK.m2.4 (joonis L2.4.4, lk 65) paistab sobivat mõnevõrra kehvemini, kuna jääkide autokorrelatsioonid 2 ja 15 on küllaltki piiripealsed. Mudeli ÜPK.m2.3 (joonis 3.4.2) korral sellist probleemi ei esine. Sellegipoolest võib lugeda mõlemad mudelid Ljung-Box testi alusel sobivaks.



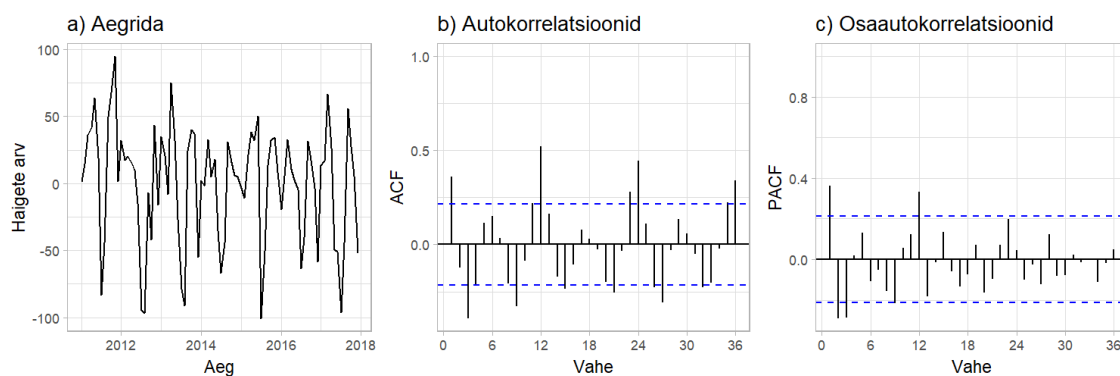
Joonis 3.4.2: Mudeli ÜPK.m2.3 jääkide diagnostika (ARIMA(1, 0, 0)(0, 1, 1)₁₂)

Mitmemõõtmelised mudelid

Kasutades regressorreana ravi rahastamise lepingu ravijuhtude arvu aegrida üldkirurgia erialal, hinnati esmalt lineaarne regressioonimudel. Selle mudeli sobitatud väärtused on toodud joonisel 3.4.3 (lk 41) ning jääkrea ülevaade joonisel 3.4.4 (lk 41).



Joonis 3.4.3: ÜPK: lineaarne regressioonimudel



Joonis 3.4.4: ÜPK: lineaarse regressioonimudeli jääkrida

Jääkaegreas esineb sesoonsus perioodiga 12, mida üritati arvesse võtta mudeli sesoonses osas AR(1) mudeliga, kuna 12 osautokorrelatsioon erineb nullist ja järgnevad sesoonsed osautokorrelatsioonid võib lugeda nulliks (ning liikuva keskmise mudel vajamiseks vähemalt 2 parameetrit). Mittesesoonses osas prooviti AR(3) ning MA(3) mudelit, kuna kolmas autokorrelatsioon ja osautokorrelatsioon on nullist erinevad, neile järgnevad aga jäävad piiridesse. Hinnatud mudelid koos grupi 24 jaoks teostatud Ljung-Box testi olulisuse tõenäosustega on toodud tabelis 3.4.2.

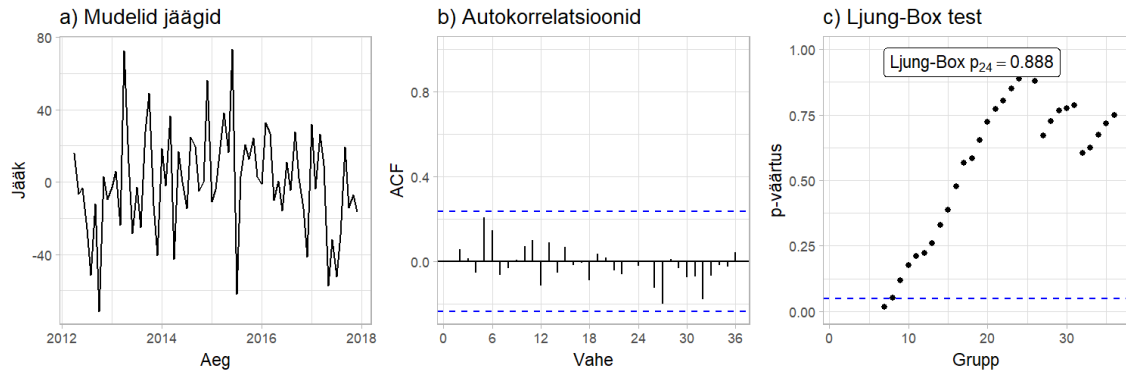
Tabel 3.4.2: ÜPK võimalikud mitmemõõtmelised mudelid

Jrk	Mudeli tähis	Mudel	$p_{LB(24)}$
1.	ÜPK.mx1.1	lineaarne regressioonimudel ARIMA(3, 0, 0)(1, 0, 0) ₁₂ vigadega	0,888 *
2.	ÜPK.mx1.2	lineaarne regressioonimudel ARIMA(0, 0, 3)(1, 0, 0) ₁₂ vigadega	0,601 *

* – mudel sobib ($p_{LB(24)} > 0,05$)

Nende mudelite jääkide aegridadel esineb mittestatsionaarsuse tunnuseid: näib, et

2012–2016 on jääkide aegreas tõusev lokaalne trend, millele järgneb langev trend (vt joonis 3.4.5, lk 42; joonis L2.4.5, lk 65). Samas on jääkide autokorrelatsioonid kuni vaheni 36 piirides ning ka Ljung-Box testi olulisuse tõenäosused suured, millest lähtuvalt loeti mõlemad mudelid ÜPK.mx1.1 ja ÜPK.mx1.2 andmetega sobivaks.



Joonis 3.4.5: Mudeli ÜPK.mx1.1 jääkide diagnostika ($\text{ARIMA}(3, 0, 0)(1, 0, 0)_{12}$)

4 Parimate mudelite valimine ja prognoosid

Antud peatükis parimaks loetud mudelite Ri väljundid on toodud lisas 4, lk 68–71.

4.1 Otorinolarüngoloogia parimad mudelid ja prognoosid

Punktis 3.1 sobivaks loetud mudelite täpsusnäitajate kokkuvõte on esitatud tabelis 4.1.1 ning aastaste prognooside ruutkeskmised vead on toodud lisas 3, joonisel L3.1 (lk 66).

Tabel 4.1.1: Ülevaade LORi mudelite täpsusest

Mudeli tähis	Treeningperioodi			$RMSE^{**}$	$RMSE^{***}$	$\hat{\sigma}$
	\overline{RMSE}^*	$RMSE_{min}$	$RMSE_{max}$	2018	2018	
LOR.m1.2	62,1	39,7(2011)	83,3(2016)	50,0	46,9	52,5
LOR.mx1.2	62,1	45,7(2017)	84,4(2016)	58,6	53,4	51,6

* – kalendriaastate 12-sammuliste prognooside $RMSE$ -de ruutkeskmise

** – 12-sammulise prognoosi ruutkeskmise viga

*** – ühesammuliste prognooside ruutkeskmise viga

Nii ühe- kui mitmemõõtmelistest sobitatud mudelitest osutus andmetega sobivaks ainult üks mudel. Ühemõõtmeline mudel LOR.m1.2 on $ARIMA(2,0,0)(2,0,0)_{12}$ eksponentsiaalse trendi eeldusel. Mitmemõõtmeline mudel on regressioonimudel $ARIMA(2,0,0)(2,0,0)_{12}$ jääkidega. Võib tähele panna, et regressioonimudeli jääkidele leitud mudel on sama tüüpi nagu trendi 1 jääkidele leitud mudel, kuna RRLi ravijuhtude arvus kajastus analoogne trend nagu uuritavas aegreas.

Ühemõõtmelise mudeli LOR.m1.2, s.o $ARIMA(2,0,0)(2,0,0)_{12}$ eksponentsiaalse trendi eeldusel, kuju on järgmine.

$$(1 - 0,255B + 0,238B^2)(1 - 0,344B^{12} - 0,532B^{24})\tilde{Z}_t = A_t,$$

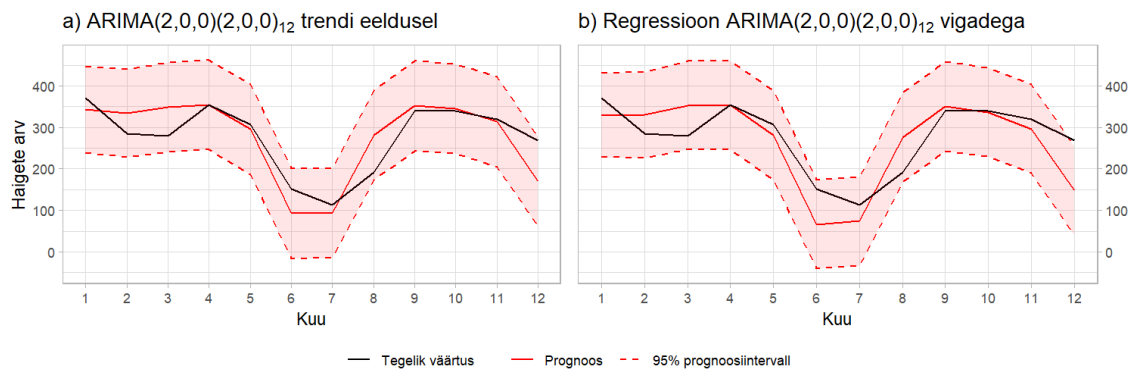
kus B on nihkeoperaator, $\tilde{Z}_t = Z_t - 235,039 - 273,336e^{c_3t}$ ning $c_3 = -0,020$ on fikseeritud vastavalt algsele eksponentsiaalse trendi mudelile.

Lõpliku mitmemõõtmelise mudeli ehk ravi rahastamise lepingu aegreast sõltuva $ARIMA(2,0,0)(2,0,0)_{12}$ vigadega lineaarse regressioonimudeli kuju on järgmine.

$$(1 - 0,191B + 0,310B^2)(1 - 0,286B^{12} - 0,636B^{24})\tilde{Z}_t = A_t,$$

kus B on nihkeoperaator, $\tilde{Z}_t = Z_t + 14,243 + \frac{0,162}{12}x_t$ ning x_t on ajahetkele t vastava aasta algul ravi rahastamise lepingus kokku lepitud LORi ravijuhtude arv.

Mudelite LOR.m1.2 ja LOR.mx1.2 prognoosid testperioodiks jaanuar–detsember 2018 koos 95%-prognoosiintervalliga on antud joonisel 4.1.1 (lk 44). Prognoosiintervallid katavad mõlema mudeli korral kogu testperioodi tegelikke väärtuseid. Samas on nii ühe- kui mitmemõõtmelise mudeli korral kahel ajahetkel prognoosiintervalli alumine piir väljaspool aegrea võimalike väärtuste piirkonda (negatiivne)⁴.



Joonis 4.1.1: Mudelite LOR.m1.2 (a) ja LOR.mx1.2 (b) prognoosid aastaks 2018

Ühe- ja mitmemõõtmelise mudeli vahel parema valimine ei ole antud juhul lihtne ülesanne. Treeningaastatel käituvad mudelid aastaste prognooside *RMSE*-de poolest küllaltki sarnaselt (võrrelda saab aastate 2014–2017 *RMSE*-sid, vt joonis L3.1, lk 66). Treeningperioodi aastate ruutkeskmised *RMSE*d on neil mudelitel võrdsed, ent nii minimaalne kui maksimaalne *RMSE* on ühemõõtmelisel mudelil mõnevõrra väiksem. Samuti on paremad ühemõõtmelise mudeli testperioodi täpsusnäitajad.

Seega võimaldab mitmemõõtmeline mudel küll vabaneda deterministliku trendi eeldusest (mis võib osutuda tugevalt veaohhtlikuks, kui tausttingimused muutuvad) ning ka taustal hinnatava eksponentsiaalse trendi parameetri kasutamisest, ent see mudel ei anna antud juhul paremaid prognoose. Mitmemõõtmelise mudeli puhul on puuduseks ka regressorrea tükiti konstantne kuju, mis viib järskude hüpeteni aasta vahetumisel.

Teatava ettevaatusega, pidades silmas selle puudusi, võib antud juhul eelistada eksponentsiaalse trendiga mudelit LOR.m1.2.

⁴Selliste probleemide lahendamiseks võib kasutada aegrea logaritmimeist, mis tagab prognooside ja prognoosiintervallide positiivsuse. Kuna antud juhul esineb aegreas langev trend ning samas taseme langusega pigem varieeruvus suureneb, siis logaritmimeine teeb rea analüüsimise keerulisemaks. Autori katsetuste põhjal ei andnud see lähenemine head tulemust.

4.2 Pediaatria parimad mudelid ja prognoosid

Punktis 3.2 sobivaks loetud mudelite täpsusnäitajate kokkuvõte on esitatud tabelis 4.2.1 ning nende mudelite aastaste prognooside ruutkeskmised vead on toodud lisas 3, joonisel L3.2 (lk 66).

Ühemõõtmelised mudelid PED.m1.1–PED.m1.4 on sobitatud diferentsimata andmetele ning need käituvad treeningperioodi aastaste prognooside $RMSE$ -de poolest sarnaselt, kusjuures mitmemõõtmelised mudelid käituvad samuti analoogselt (joonis L3.2, lk 66). Nimetatud mudelite treeningperioodi aastaste prognooside ruutkeskmised $RMSE$ -d on ligikaudu võrdsed, minimaalsed $RMSE$ -d on samuti lähedased, ehkki saavutatakse ühemõõtmeliste mudelite korral aastal 2012, mitmemõõtmeliste mudelite korral aastal 2014 (viimaste abil aasta 2012 prognoosi arvutada ei ole võimalik).

Tabel 4.2.1: Ülevaade pediaatria mudelite täpsusest

Mudeli tähis	Treeningperioodi			$RMSE^{**}$	$RMSE^{***}$	$\hat{\sigma}$
	\overline{RMSE}^*	$RMSE_{min}$	$RMSE_{max}$	2018	2018	
PED.m1.1	30,4	22,7 (2012)	40,0 (2013)	25,6	23,8	27,2
PED.m1.2	30,5	22,5 (2012)	40,4 (2013)	26,6	24,8	27,3
PED.m1.3	30,2	22,9 (2012)	39,6 (2013)	24,0	21,8	26,8
PED.m1.4	30,3	22,7 (2012)	39,9 (2013)	24,7	22,4	26,9
PED.m2.2	28,1	20,8 (2014)	39,5 (2013)	18,7	24,8	28,7
PED.mx1.1	29,5	22,5 (2015)	36,8 (2013)	29,4	26,3	25,8
PED.mx1.2	29,5	22,8 (2015)	36,9 (2013)	29,8	26,7	25,9
PED.mx1.3	29,5	22,5 (2015)	37,2 (2013)	28,2	24,5	25,8
PED.mx1.4	29,6	23,0 (2015)	37,7 (2013)	28,9	25,1	25,9

* – kalendriaastate 12-sammuliste prognooside $RMSE$ -de ruutkeskmise

** – 12-sammulise prognoosi ruutkeskmise viga

*** – ühesammuliste prognooside ruutkeskmise viga

Kõik vaadeldud mudelid prognoosivad ühtviisi halvasti aastat 2013, mille prognoosi ruutkeskmise viga on teiste aastatega võrreldes kõigi mudelite korral suurim. Antud aasta oli ka vaadeldud aastatest kõige suurema haigete arvuga ehk mingis mõttes teistest natuke erinev.

Kõigist mudelitest eristub mõnevõrra mudel PED.m2.2, mis eeldab, et sesoonsus aegreas on jääv ehk mudelis on aegrida sesoonselt diferentsitud. See mudel on teistega võrreldes täpsem nii treeningperioodi aastaste prognooside ruutkeskmise $RMSE$ kui minimaalse $RMSE$ poolest ning test aasta 2018 12-sammulise prognoosi ruutkeskmise vea poolest. Seetõttu võiks selle mudeli lugeda antud töös parimaks pediaatria ühemõõtmeliseks mudeliks.

Valitud lõplik ühemõõtmeline mudel PED.m2.2 on $ARIMA(1, 0, 0)(0, 1, 1)_{12}$ mudel, mille kuju on järgmine.

$$(1 - 0,234B)(1 - B^{12})Z_t = (1 - 0,622B^{12})A_t,$$

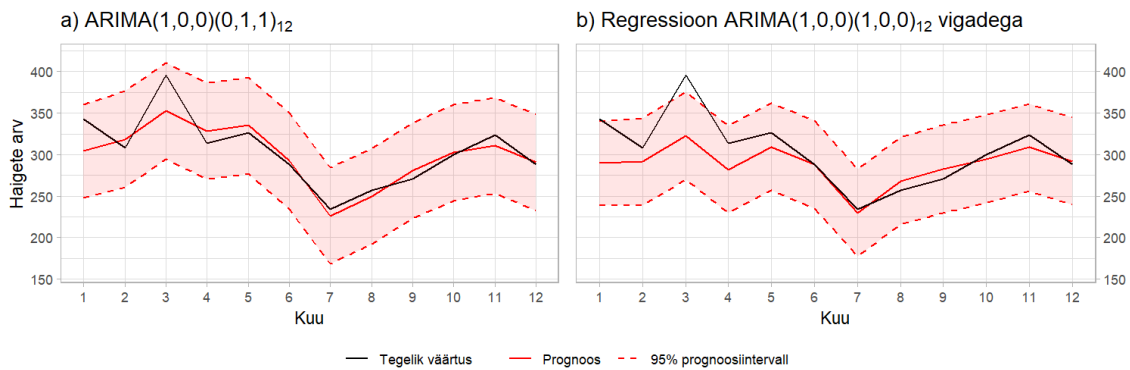
kus B on nihkeoperaator. Mudeli kuju peale mõningaid teisendusi on

$$Z_t = 0,234Z_{t-1} + Z_{t-12} - 0,234Z_{t-13} + A_t - 0,622A_{t-12}.$$

Nagu öeldud, on mitmemõõtmelised mudelid täpsusnäitajate poolest sarnased. Teistest veidi täpsemad on PED.mx1.1 ja PED.mx1.3. Parima valikul lähtuti mudelis hinnatud parameetrite arvust, mistõttu loeti parimaks PED.mx1.1, s.o RRLi ravijuhtude arvust sõltuv regressioonmudel $ARIMA(1, 0, 0)(1, 0, 0)_{12}$ vigadega, kus on hinnatud üks parameeter vähem kui mudelis PED.mx1.3. Valitud mudeli kuju on järgmine.

$$(1 - 0,286B)(1 - 0,626B^{12})\tilde{Z}_t = A_t,$$

kus B on nihkeoperaator, $\tilde{Z}_t = Z_t - 396,107 + \frac{0,510}{12}x_t$ ning x_t on ajahetkele t vastava aasta algul ravi rahastamise lepingus kokku lepitud pediaatria ravijuhtude arv.



Joonis 4.2.1: Mudelite PED.m2.2 (a) ja PED.mx1.1 (b) prognoosid aastaks 2018

Mudelite PED.m2.2 ning PED.mx1.1 prognoosid testperioodiks jaanuar–detsember 2018 koos 95%-prognoosiintervallidega on antud joonisel 4.2.1. Ühemõõtmelise mudeli korral prognoosiintervall katab kõiki testperioodi tegelikke väärtusi. Mitmemõõtmelise korral aga on jaanuaris ja märtsis 2018 tegelik haigete arv veidi üle prognoosiintervalli ülemise piiri. Ühe- ja mitmemõõtmelise mudeli vahel valides võiks seetõttu antud juhul eelistada ühemõõtmelist mudelit PED.m2.2.

4.3 Sisehaiguste parimad mudelid ja prognoosid

Punktis 3.3 sobivaks loetud mudelite täpsusnäitajate kokkuvõte on esitatud tabelis 4.3.1 ning nende mudelite aastaste prognooside ruutkeskmised vead on toodud lisas 3, joonisel L3.3 (lk 66).

Ühemõõtmeliste mudelite täpsusnäitajad annavad kinnitust töös varasemalt püstitatud hüpoteesile, et trendi 2 (s.o vabalt muutuv trendikõver kummalgi pool muutuse perioodi) eeldusel mudeli loomine antud andmete korral ei vii headele tulemustele. Mudeli SISE.m2.3 ehk trendi 2 eeldusel saadud mudeli täpsusnäitajad viitavad ülesobitamisele. Ehkki treeningperioodis on mudeli SISE.m2.3 ruutkeskmised vead valdavalt väiksemad kui mudelil SISE.m1.2, on testperioodis nii 12-sammulise prognoosi kui ühesammuliste prognooside vead märkimisväärselt suuremad kui mudelil SISE.m1.2. Samuti on mudeli m2.3 testperioodi ühesammuliste prognooside ruutkeskmine viga märkimisväärselt suurem mudeli standardveast, mis viitab samuti ülesobitamisele. Võrreldes testperioodi prognoose tegelike väärtustega (vt joonis L2.3.5, lk 63), on näha, et mudel SISE.m2.3 annab ülehinnangud, kusjuures alates maist 2018 ei kata ka 95%-prognoosiintervallid tegelikke väärtusi. Seega mudeli SISE.m2.3 näol ei ole tegu hea prognoosivõimega mudeliga. Samas on just trendi 2 tüüpi trendi eeldus taustinfole tuginedes korrektsem, mistõttu võiks andmete lisandumisel tulevikus sellist mudelit uuesti proovida.

Tabel 4.3.1: Ülevaade sisehaiguste mudelite täpsusest

Mudeli tähis	Treeningperioodi			$RMSE^{**}$	$RMSE^{***}$	$\hat{\sigma}$
	\overline{RMSE}^*	$RMSE_{min}$	$RMSE_{max}$	2018	2018	
SISE.m1.2	20,1	13,1 (2014)	30,3 (2016)	34,6	23,8	17,4
SISE.m2.3	17,8	11,8 (2014)	24,4 (2017)	58,8	44,8	16,4
SISE.mx1.1	20,8	11,9 (2014)	29,7 (2017)	25,9	23,8	17,4
SISE.mx1.2	19,9	11,8 (2014)	29,1 (2017)	26,3	24,8	17,7

* – kalendriaastate 12-sammuliste prognooside $RMSE$ -de ruutkeskmine

** – 12-sammulise prognoosi ruutkeskmine viga

*** – ühesammuliste prognooside ruutkeskmine viga

Parimaks ühemõõtmeliseks mudeliks loeti eelneva põhjal mudel SISE.m1.2 ehk $ARIMA(3, 0, 0)(2, 0, 0)_6$ trendi 1 eeldusel (s.o trend erineva taseme, aga ühesuguse tõusuga kummalgi pool muutuse perioodi). Selle mudeli kuju on järgmine.

$$(1 - 0,251B - 0,222B^2 - 0,026B^3)(1 - 0,237B^6 - 0,512B^{12})\tilde{Z}_t = A_t,$$

kus B on nihkeoperaator, $\tilde{Z}_t = Z_t - 135,569 - 1,387t + 10,559g(t)$ ning

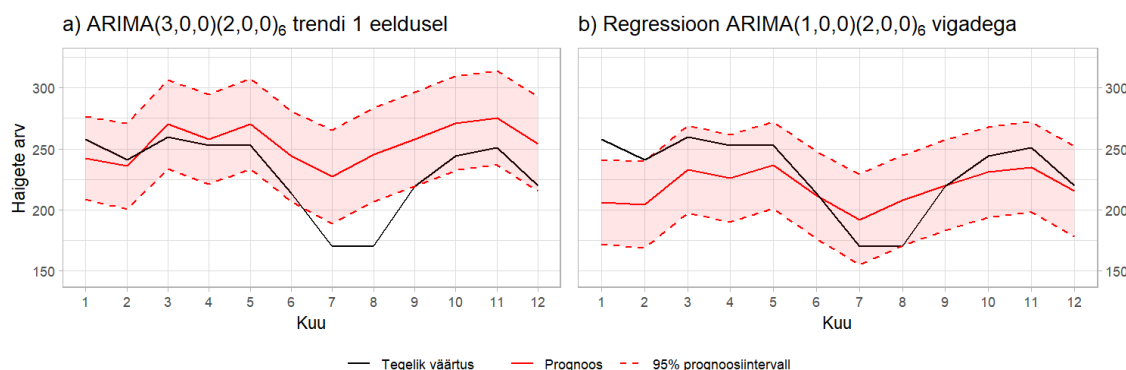
$$g(t) = \begin{cases} 0, & t \leq 97 \text{ (jaanuar 2016)} \\ t - 97, & 97 < t < 103 \\ 6, & t \geq 103 \text{ (juuli 2016)} \end{cases}.$$

Mitmemõõtmelised mudelid on täpsusnäitajate poolest küllaltki sarnased. Treeningperioodis prognoosib mõnevõrra paremini mudel SISE.mx1.2 ehk regressioonimudel ARIMA(0,0,1)(2,0,0)₆ vigadega. Testperioodis on aga täpsem mudel SISE.mx1.1 ehk regressioonimudel ARIMA(1,0,0)(2,0,0)₆ vigadega. Kuna mudel SISE.mx1.1 on mõnevõrra lihtsama kujuga (sisaldab ainult autoregressiivseid komponente), siis valiti see parimaks modeliks.

Mudeli SISE.mx1.1 mudeli kuju on järgmine.

$$(1 - 0,293B)(1 - 0,284B^6 - 0,432B^{12})\tilde{Z}_t = A_t,$$

kus B on nihkeoperaator, $\tilde{Z}_t = Z_t + 243,666 - \frac{2,062}{12}x_t$ ning x_t on ajahetkele t vastava aasta algul ravi rahastamise lepingus kokku lepitud sisehaiguste katuseriala alla kuuluvate erialade ravijuhtude koguarv.



Joonis 4.3.1: Mudelite SISE.m1.2 (a) ja SISE.mx1.1 (b) prognoosid aastaks 2018

Mudelite SISE.m1.2 ja SISE.mx1.1 prognoosid testperioodiks jaanuar–detsember 2018 koos 95%-prognoosiintervalliga on toodud joonisel 4.3.1. Ei ühe- ega ka mitmemõõtmelise mudeli prognoosiintervall ei kata kõiki testperioodi tegelikke väärtusi. Ühemõõtmelise mudeli korral jäävad prognoosiintervallist välja juuli ja augusti tegelikud väärtused, mis on prognoositust madalamad. Mitmemõõtmeline mudel aga alahindab jaanuari ja veebruari väärtusi, mis jäävad prognoosiintervallist välja.

Ühe- ja mitmemõõtmelise mudeli võrdluses on antud juhul mõistlik eelistada

mitmemõõtmelist mudelit SISE.mx1.1, kuna selle minimaalne ja maksimaalne treeningperioodi $RMSE$ on väiksemad, testperioodi 12-sammulise prognoosi $RMSE$ on samuti väiksem ning selles mudelis on hinnatud vaid 5 parameetrit, samal ajal ühemõõtmelises mudelis on hinnatud 8. Ühtlasi võimaldab mitmemõõtmelise mudeli kasutamine vältida deterministliku trendi sobitamist.

Seejuures ei ole ka valitud mitmemõõtmeline mudel täielikult probleemideta. Nagu eelnevalt mainitud, toob regressorrea tükiti konstantne kuju kaasa ebatäpsusi eriti iga uue aasta esimeste kuude prognoosides, kusjuures praktikas selgub täpne uue aasta RRLi tellitav ravijuhtude arv alles esimese kvartali jooksul, mitte päris aasta alguses (kirjalikult lepitakse kokku hiljemalt jooksva aasta 15. veebruariks [28]). Selle hetkeni toimub ravi eest tasumine hoopis eelneva aasta lepingu põhjal [28]. Nende asjaoludega tuleb mudeli kasutamisel arvestada.

4.4 Üld- ja plastikakirurgia parimad mudelid ja prognoosid

Punktis 3.4 sobivaks loetud mudelite täpsusnäitajate kokkuvõte on esitatud tabelis 4.4.1 ning nende mudelite aastaste prognooside ruutkeskmised vead on toodud lisas 3, joonisel L3.4 (lk 67).

Tabel 4.4.1: Ülevaade ÜPK mudelite täpsusest

Mudeli tähis	Treeningperioodi			$RMSE^{**}$	$RMSE^{***}$	$\hat{\sigma}$
	\overline{RMSE}^*	$RMSE_{min}$	$RMSE_{max}$	2018	2018	
ÜPK.m1.5	30,8	19,5 (2014)	50,1 (2011)	19,0	17,2	25,0
ÜPK.m2.3	29,1	23,0 (2014)	38,5 (2017)	17,0	16,2	27,0
ÜPK.m2.4	29,0	24,0 (2014)	37,9 (2017)	18,2	16,7	28,0
ÜPK.mx1.1	31,1	23,9 (2014)	35,6 (2013)	30,8	27,9	27,2
ÜPK.mx1.2	30,9	23,4 (2014)	34,9 (2017)	31,3	28,4	28,2

* – kalendriaastate 12-sammuliste prognooside $RMSE$ -de ruutkeskmise

** – 12-sammulise prognoosi ruutkeskmise viga

*** – ühesammuliste prognooside ruutkeskmise viga

Jooniselt L3.4 (lk 67) on näha, et ühemõõtmelised mudelid ÜPK.m2.3–ÜPK.m2.4 sesoonse diferentsiga käituvad aastaste prognooside täpsuse mõttes ühtmoodi, mitmemõõtmelised mudelid ÜPK.mx1.1–ÜPK.mx1.2 teisiti ning kõigist neist eristub ühemõõtmeline mudel ÜPK.m1.5 diferentsimata andmetel. Viimane on aastal 2011 selgelt teistest ühemõõtmelistest mudelitest ebatäpsem (mistõttu on ka treeningperioodi aastaste prognooside ruutkeskmise $RMSE$ nendega võrreldes suurem). Kogu järgnevas perioodis on see mudel aga pigem sama täpne või täpsem, kuid on ka oma kujult keerukam, sest see eeldab sesoonses osas ARMA(1,1) mudelit. Testperioodis aga see keerukus eelist ei anna: 2018. aasta

prognooside täpsuse poolest on siiski ÜPK.m2.3 parim nii ühe- kui mitmemõõtmeliste mudelitega võrreldes.

Eelneva arutelu põhjal loeti parimaks ühemõõtmeliseks mudeliks ÜPK.m2.3 ehk mudel $ARIMA(1, 0, 0)(0, 1, 1)_{12}$. Valitud mudeli kuju on järgmine.

$$(1 - 0,350B)(1 - B^{12})Z_t = (1 - 0,770B^{12})A_t,$$

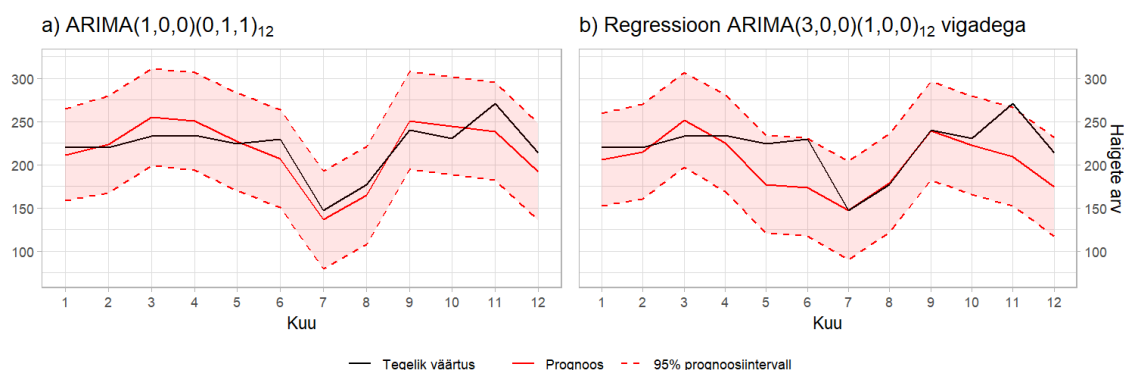
kus B on nihkeoperaator. Peale teisendusi omandab mudel kuju

$$Z_t = 0,350Z_{t-1} + Z_{t-12} - 0,350Z_{t-13} + A_t - 0,770A_{t-12}.$$

Parimaks mitmemõõtmeliseks mudeliks loeti mudel ÜPK.mx1.1, kuna treeningperioodis vaadeldud 2 mudelit oluliselt ei erinenud, testperioodi täpsusnäitajate poolest oli ÜPK.mx.1 mõnevõrra parem ning see mudel on ka lihtsama kujuga (sisaldab ainult autoregressiivseid liikmeid). Mudeli ÜPK.mx1.1 ehk $ARIMA(3, 0, 0)(1, 0, 0)_{12}$ vigadega lineaarse regressioonimudeli kuju on järgmine.

$$(1 - 0,224B + 0,045B^2 + 0,252B^3)(1 - 0,596B^{12})\tilde{Z}_t = A_t,$$

kus B on nihkeoperaator, $\tilde{Z}_t = Z_t - 177,168 - \frac{0,132}{12}x_t$ ning x_t on ajahetkele t vastava aasta algul ravi rahastamise lepingus kokku lepitud ÜPK ravijuhtude arv.



Joonis 4.4.1: Mudelite ÜPK.m2.3 (a) ja ÜPK.mx1.1 (b) prognoosid aastaks 2018

Mudelite ÜPK.m2.3 ja ÜPK.mx1.1 prognoosid testperioodiks jaanuar–detsember 2018 koos 95%-prognoosiintervalliga on toodud joonisel 4.4.1. Ühemõõtmelise mudeli korral katavad prognoosiintervallid kogu testperioodi tegelikke väärtusi, mitmemõõtmelise mudeli korral on aga 2018. aasta novembris tegelik väärtus veidi prognoosiintervalli ülemisest piirist suurem. Ühe- ja mitmemõõtmelise mudeli võrdluses on antud juhul eelistatum ühemõõtmeline mudel ÜPK.m2.3, kuna selle prognoosivõime on testperioodis märgatavalt parem.

Kokkuvõte

Käesoleva magistritöö eesmärgiks oli prognoosida Tartu Ülikooli Kliinikumi haiglapatsientide arvu kuus neljal ravierialal kuni üks aasta ette. Seejuures keskenduti sobivate ühemõõtmeliste ARIMA tüüpi prognoosimudelite leidmisele. Proovimaks saada täpsemaid prognoose, rakendati ka lineaarset regressiooni ARIMA tüüpi vigadega, kus regressorina kasutati Eesti Haigekassa poolt aastaks tellitavate ravijuhtude arve vastavatel erialadel.

Uurimise all olid järgmised erialad: otorinolarüngoloogia, pediaatria, sisehaigused ning üld- ja plastikakirurgia. Iga eriala jaoks valiti parim ühe- ja mitmemõõtmeline mudel ning võrreldi nende täpsust aastaste prognooside ruutkeskmiste vigade põhjal. Saadud mudelite abil leiti 12-sammulised prognoosid valimiväliseks aastaks 2018 koos prognoosiintervallidega ning võrreldi tulemusi selle aasta tegelike ravitud haigete arvudega.

Antud töö põhitulemus seisneb selles, et vähemalt haiglaravi patsientide arvu prognoosimiseks vaadeldud erialadel on võimalik leida ARIMA tüüpi mudelite klassist nii ühe- kui mitmemõõtmelisi andmetega sobivaid mudeleid. Samas tuleb mitmete saadud mudelite kasutamisel olla ettevaatlik, kuna nende korrektsuse või täpsuse osas esineb küsitavusi.

Esiteks on antud töö kontekstis parimaks tunnistatud ühemõõtmeliste mudelite seas deterministliku trendi eeldusel saadud mudeleid, mis on korrektsed vaid juhul, kui sellise trendi eeldamine on mõistlik. Juhul kui sellise trendi eeldamine ei ole reaalsusega kooskõlas või olemasolevad andmed ei võimalda seda trendi hästi hinnata, võivad kaugemasse tulevikku tehtavad prognoosid olla väga ebatäpsed.

Teiseks on otorinolarüngoloogia eriala jaoks leitud mudelite korral testperioodis mõne ajapunkti prognoosiintervalli alumine piir negatiivne. Samas on ilmne, et haigete arv ei saa olla negatiivne. See on näide juhust, kus pigem suure varieeruvuse ning mõningate väiksemate aegrea väärtuste esinemise tõttu ilmneb, et loendusandmete korral ei pruugi normaaljaotuse eeldusel arvutatud prognoosiintervallid, aga ka ARIMA tüüpi mudelid üldiselt hästi sobida. Kui vaatluse all oleks kitsamad ravierialad või ka päevaravi, kus haigete arvud on pigem väikesed ning võivad langeda mõnes kuus ka nulli, siis oleks arvatavasti õigem kasutada mudeleid, mis eeldavad nt loendusandmete omast Poissoni jaotust. Antud töö edasiarendusena võiks uurida, kuidas sellised mudelid töötaks vaadeldavate erialade korral ning võrrelda tulemusi. Loendusandmete aegridade mudelite hindamise võimalused on realiseeritud näiteks rakendustarkvara R pakettides *glarma*, *acp*, *tscout* [29], [30], [31].

Ka antud töös vaadeldud mitmemõõtmeliste mudelite kasutamisel tuleks olla tähelepanelik nende kitsaskohtade suhtes. Kuna uuritavad aegread on kuu tasandil ning huvipakkuvad regressorid aastase intervalliga, siis teisendati viimased töös kasutamiseks kuu tasandile tükiti konstantsena. Selline käsitus toob kaasa järsud muutused regressorreas iga aasta algul, ehkki ravitavate haigete arvu päris esimestel kuudel uue aasta leping tegelikult ei mõjuta [28] ning ka järgnevad muutused haigete arvus ilmselt nii järsult ei toimu. Seetõttu võib esineda probleeme prognooside täpsuses eriti aasta algul. Magistritööd võiks edasi arendada täiendavate andmete lisandumisel, kasutades regressori ja uuritava aegrea seose uurimiseks näiteks hoopis aasta tasandil ridu ning püüda seejärel leitud seost kanda üle kuu tasandile prognooside täpsustamiseks. Alternatiivselt võiks proovida mõelda välja viisi, kuidas teisendada aastast ravijuhtude arvu kuu tasandile mingil siledamal meetodil. Kuna antud töös olid kasutada vaid 8 aasta pikkused regressorread, siis neid lähenemisi ei proovitud.

Viidatud allikad

- [1] *Sihtasutuse Tartu Ülikooli Kliinikumi põhikiri*. 12. aprill 2018. URL: <https://www.kliinikum.ee/et/alusdokumendid/pohikiri> (vaadatud 03.03.2019).
- [2] *Konsolideeritud majandusaasta aruanne*. Sihtasutus Tartu Ülikooli Kliinikum, 2018. URL: https://www.kliinikum.ee/pildid/tutvustus/tegevusaruanded/Sihtasutuse_Tartu_Ylikooli_Kliinikum_2018a_tgevusaruanne.pdf (vaadatud 02.05.2019).
- [3] *Tervishoiu arengusuunad aastani 2020*. Sotsiaalministeerium. URL: https://www.sm.ee/sites/default/files/content-editors/eesmargid_ja_tgevused/Tervis/tervishoiu_arengusuunad_2020.pdf (vaadatud 03.03.2019).
- [4] Kangro, R. „Aegridade analüüs“ (2012). URL: <https://dspace.ut.ee/handle/10062/27703> (vaadatud 03.03.2019).
- [5] Shumway, R. H. ja Stoffer, D. S. *Time Series Analysis and Its Applications: With R Examples*. 4. väljaanne. Springer Series in Statistics. Springer International Publishing, 2017.
- [6] Käärrik, M. ja Pärna, K. „Juhuslikud protsessid“ (2013). URL: <https://dspace.ut.ee/handle/10062/29478> (vaadatud 03.03.2019).
- [7] Hyndman, R. J. ja Athanasopoulos, G. *Forecasting: Principles and Practice*. 2. väljaanne. Melbourne, Australia: OTexts, 2018. URL: <https://otexts.com/fpp2/> (vaadatud 17.02.2019).
- [8] Box, G. E. P. ja Jenkins, G. M. *Time Series Analysis: Forecasting and Control*. Holden-Day Series in Time Series Analysis. Holden-Day, Inc, 1970.
- [9] Ljung, G. M. ja Box, G. E. P. „On a Measure of Lack of Fit in Time Series Models“. *Biometrika* 65 (august 1978). DOI: 10.1093/biomet/65.2.297.
- [10] *Eesti tervisestatistika raamat 2005-2008*. Sotsiaalministeerium, 2010. URL: <http://www.digar.ee/id/nlib-digar:272475>.
- [11] *Tervishoiuteenuste korraldamise seadus*. 2019. URL: <https://www.riigiteataja.ee/akt/101022019013> (vaadatud 03.03.2019).
- [12] Eesti Haigekassa. *Ravi rahastamise lepingu lisatingimused eriarstiabis, hambaravis ja õendusabis (ravi rahastamise lepingu lisa 2)*. 16. jaanuar 2019. URL: https://www.haigekassa.ee/sites/default/files/RRL%202019/RRL_lisa_2_2019_30.pdf (vaadatud 15.03.2019).
- [13] Eesti Haigekassa. *Tervishoiusüsteemi korraldus Eestis*. URL: <https://www.haigekassa.ee/kontaktpunkt/tervishoiususteemi-korraldus-eestis> (vaadatud 03.03.2019).

- [14] *Konsolideeritud majandusaasta aruanne*. Sihtasutus Tartu Ülikooli Kliinikum, 2017. URL: https://www.kliinikum.ee/pildid/tutvustus/tegevusaruanded/Sihtasutuse_Tartu_Ylikooli_Kliinikum_2017a_tegevusaruanne.pdf (vaadatud 03.03.2019).
- [15] Eesti Haigekassa. *Tervishoiuteenuste nõudluse hindamine*. URL: <https://www.haigekassa.ee/partnerile/raviasutusele/ravi-rahastamise-lepingud/tervishoiuteenuste-noudluse-hindamine> (vaadatud 15.03.2019).
- [16] Eesti Haigekassa. *Tervishoiuteenuste nõudluse hindamine 2018. aastaks*. URL: https://www.haigekassa.ee/sites/default/files/n%C3%B5udlus/07.hinnatud_noudlus.pdf (vaadatud 15.03.2019).
- [17] SA Tartu Ülikooli Kliinikum. *TKL-01. Struktuuriüksuste koodid*. Juhtimissüsteemi dokument. Versioon 166. 18. detsember 2018.
- [18] *Otorinolarüngoloogia eriala arengukava aastani 2020*. Sotsiaalministeerium, 2012. URL: https://www.sm.ee/sites/default/files/content-editors/eesmargid_ja_tegevused/Tervis/Tervishoiususteem/Arstide_erialade_arengukavad/otorinolaryngoloogia_arengukava_16_01_2014_.pdf (vaadatud 15.03.2019).
- [19] *Pediaatria eriala arengukava aastani 2020*. Sotsiaalministeerium, 2012. URL: https://www.sm.ee/sites/default/files/content-editors/eesmargid_ja_tegevused/Tervis/Tervishoiususteem/Arstide_erialade_arengukavad/pediaatria_arengukava.pdf (vaadatud 15.03.2019).
- [20] Eesti Sisearstide Ühendus, Lember, M., Lill, A., Nõukas, K., Saadla, P., Sukles, K. ja Toomik, M. *Sisehaiguste eriala arengukava aastani 2020*. Sotsiaalministeerium, 2012. URL: https://www.sm.ee/sites/default/files/content-editors/eesmargid_ja_tegevused/Tervis/Tervishoiususteem/Arstide_erialade_arengukavad/sisehaigused_arengukava.pdf (vaadatud 15.03.2019).
- [21] Salupere, R., Margus, B., Kariis, T., Rimmel, T., Peetsalu, M. ja Poliitikauuringute Keskus Praxis. *Gastroenteroloogia eriala arengukava aastani 2020*. Sotsiaalministeerium, 2012. URL: https://www.sm.ee/sites/default/files/content-editors/eesmargid_ja_tegevused/Tervis/Tervishoiususteem/Arstide_erialade_arengukavad/gastroenteroloogia_arengukava.pdf (vaadatud 15.03.2019).

- [22] Jakovlev, Ü., Kelk, E., Lubi, M., Rajasalu, T. ja Volke, V. *Endokrinoloogia eriala arengukava aastani 2020*. Sotsiaalministeerium, 2012. URL: https://www.sm.ee/sites/default/files/content-editors/eesmargid_ja_tegevused/Tervis/Tervishoiususteem/Arstide_erialade_arengukavad/endokrinoloogia_arengukava.pdf (vaadatud 15.03.2019).
- [23] *Reumatoloogia eriala arengukava aastani 2020*. Sotsiaalministeerium, 2012. URL: https://www.sm.ee/sites/default/files/content-editors/eesmargid_ja_tegevused/Tervis/Tervishoiususteem/Arstide_erialade_arengukavad/reumatoloogia_arengukava.pdf (vaadatud 15.03.2019).
- [24] *Konsolideeritud majandusaasta aruanne*. Sihtasutus Tartu Ülikooli Kliinikum, 2016. URL: https://www.kliinikum.ee/pildid/tutvustus/tegevusaruanded/Sihtasutuse_Tartu_Ylikooli_Kliinikum_2016a_tegevusaruanne.pdf (vaadatud 03.03.2019).
- [25] *Plastika- ja rekonstruktiivkirurgia eriala arengukava aastani 2020*. Sotsiaalministeerium, 2012. URL: https://www.sm.ee/sites/default/files/content-editors/eesmargid_ja_tegevused/Tervis/Tervishoiususteem/Arstide_erialade_arengukavad/plastika_ja_rekonstruktiivkirurgia_arengukava.pdf (vaadatud 15.03.2019).
- [26] Üldkirurgide Selts, Teras, J., Lepner, U., Arak, A., Tepp, J. ja Rander, R. *Üldkirurgia eriala arengukava aastani 2020*. Sotsiaalministeerium, 2012. URL: https://www.sm.ee/sites/default/files/content-editors/eesmargid_ja_tegevused/Tervis/Tervishoiususteem/Arstide_erialade_arengukavad/uldkirurgia_arengukava.pdf (vaadatud 15.03.2019).
- [27] Eesti Haigekassa. *Ravi rahastamise lepingud*. URL: <https://www.haigekassa.ee/partnerile/raviasutusele/ravi-rahastamise-lepingud> (vaadatud 15.03.2019).
- [28] Eesti Haigekassa. *Ravi rahastamise lepingu üldtingimused (ravi rahastamise lepingu lisa 1)*. 20. detsember 2018. URL: https://www.haigekassa.ee/sites/default/files/RRL%202019/RRL_lisa_1_2018_674.pdf (vaadatud 15.03.2019).
- [29] Dunsmuir, W. T. M., Li, C. ja Scott, D. J. *Package 'glarma'. Generalized Linear Autoregressive Moving Average Models*. 7. veebruar 2018. URL: <https://cran.r-project.org/package=glarma> (vaadatud 15.03.2019).

- [30] Vasileios, S. *Package 'acp'. Autoregressive Conditional Poisson*. 4. detsember 2015. URL: <https://cran.r-project.org/package=acp> (vaadatud 15.03.2019).
- [31] Liboschik, T., Fried, R., Fokianos, K., Probst, P. ja Rathjens, J. *Package 'tscount'. Analysis of Count Time Series*. 24. november 2017. URL: <https://cran.r-project.org/package=tscount> (vaadatud 15.03.2019).

Lisad

Lisa 1 Katuserialade seosed arstlike erialade ja kliinikumi allüksustega

Tabel L1.1: Katuserialade seosed arstlike erialade ja kliinikumi allüksustega

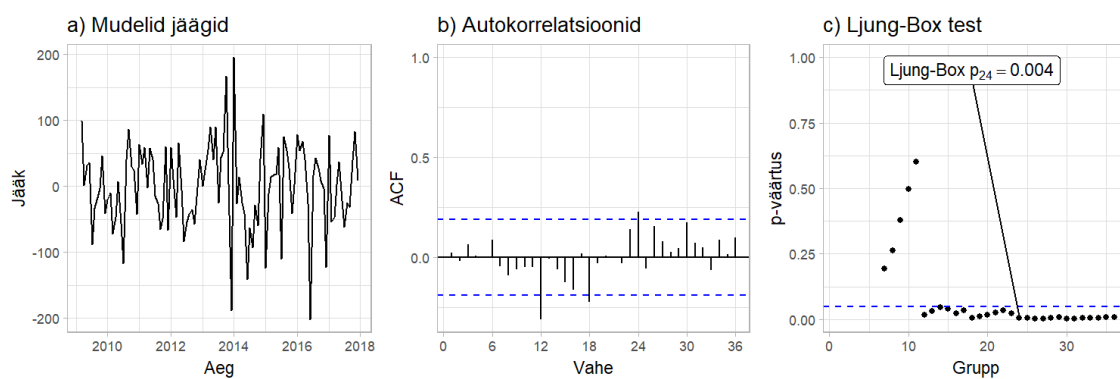
Katuseriala	Arstlikud erialad*	Kliinikumi struktuuriüksused, kood ja nimetus (s – suletud, kehtivus seisuga 31.12.2018)
LOR	otorino-larüngoloogia	461214 kõrvakliiniku statsionaar 461010 kõrva- nina- kurguhaiguste statsionaar (s) 461017 kõrva-nina-kurguhaiguste statsionaar Pärnus (s) 461111 kuulmise ja kõnestamise statsionaar (s)
Pediaatria	pediaatria	471013 neonatoloogia statsionaar 471711 üldpediaatria statsionaarne 471714 neuroloogia statsionaarne 471210 ägedate infektsioonide statsionaar 471010 lastekliiniku neonatoloogia (s) 471110 neuroloogia ja neuror rehabilitatsiooni statsionaar (s) 471310 üldpediaatria statsionaar (s)
Sisehaigused	sisehaigused, gastro-enteroloogia, reumatoloogia, endokrinoloogia	532010 gastroenteroloogia statsionaar 532110 sisehaiguste statsionaar 532111 reumatoloogia statsionaar 532112 endokrinoloogia statsionaar 531710 endokrinoloogia statsionaar (s) 531712 gastroenteroloogia statsionaar (s) 531810 reumatoloogia statsionaar (s) 531811 sisehaiguste statsionaar (s)
ÜPK	üldkirurgia, plastikakirurgia**	441010 abdominaalkirurgia statsionaar 441310 üldkirurgia statsionaar 441311 plastilise kirurgia statsionaar

* - Arvestatud samanimelisi ridu RRLis.

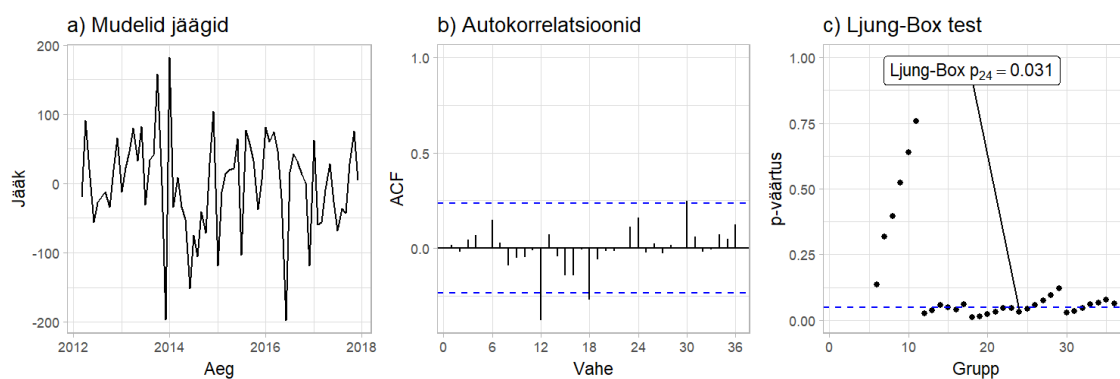
** - Plastikakirurgia eriala RRLis eraldi ei kajastu.

Lisa 2 Täiendavaid mudelite jooniseid

Lisa 2.1 Otorinolarüngoloogia täiendavaid mudelite jooniseid

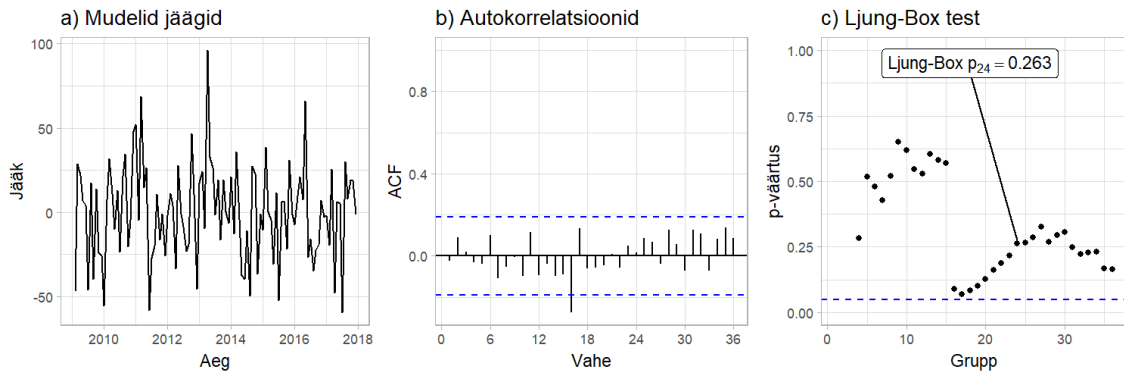


Joonis L2.1.1: Mudeli LOR.m1.1 jääkide diagnostika ($ARIMA(2, 0, 0)(1, 0, 0)_{12}$ eksponentsiaalse trendi eeldusel)

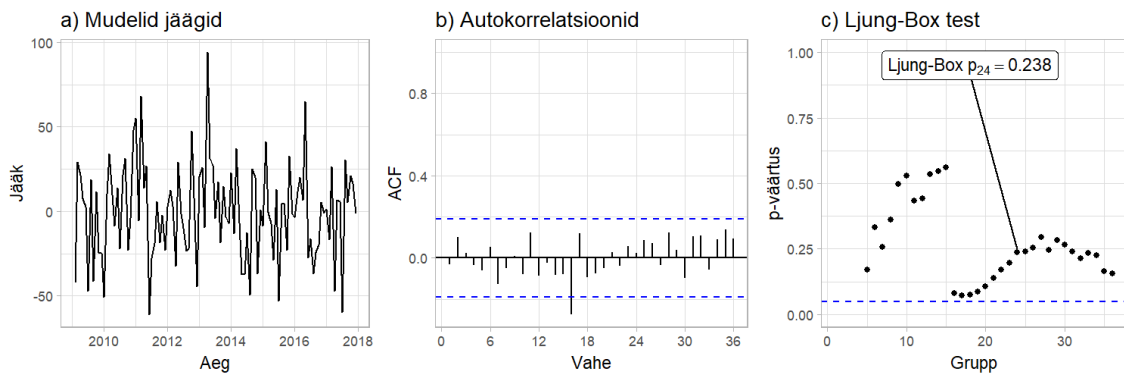


Joonis L2.1.2: Mudeli LOR.mx1.1 jääkide diagnostika (lineaarne regressioonimudel $ARIMA(2, 0, 0)(1, 0, 0)_{12}$ vigadega)

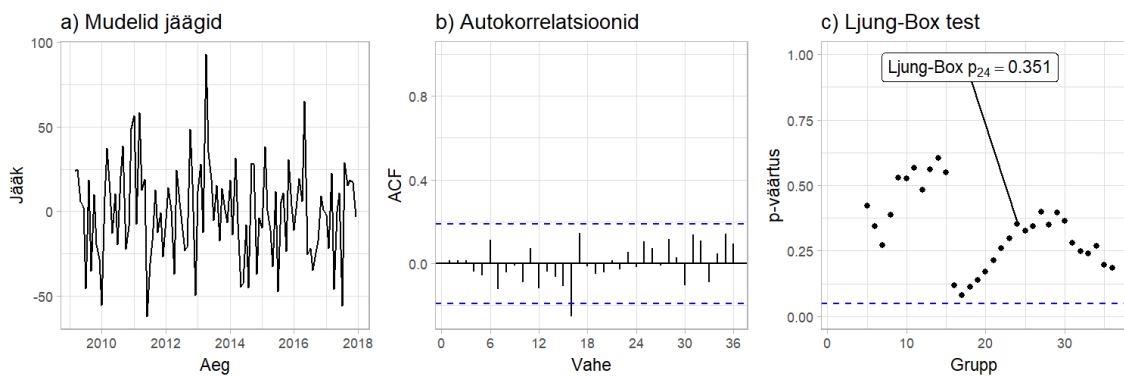
Lisa 2.2 Pediaatria täiendavaid mudelite jooniseid



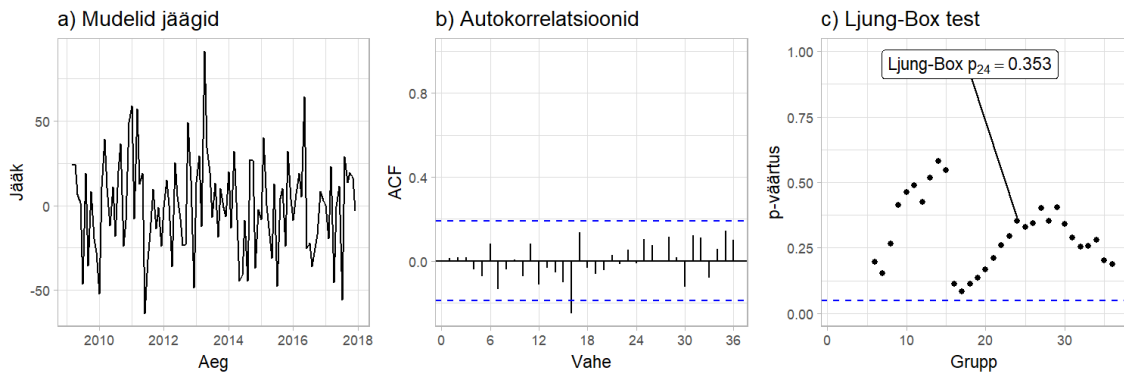
Joonis L2.2.1: Mudeli PED.m1.1 jääkide diagnostika (ARIMA(1,0,0)(1,0,0)₁₂)



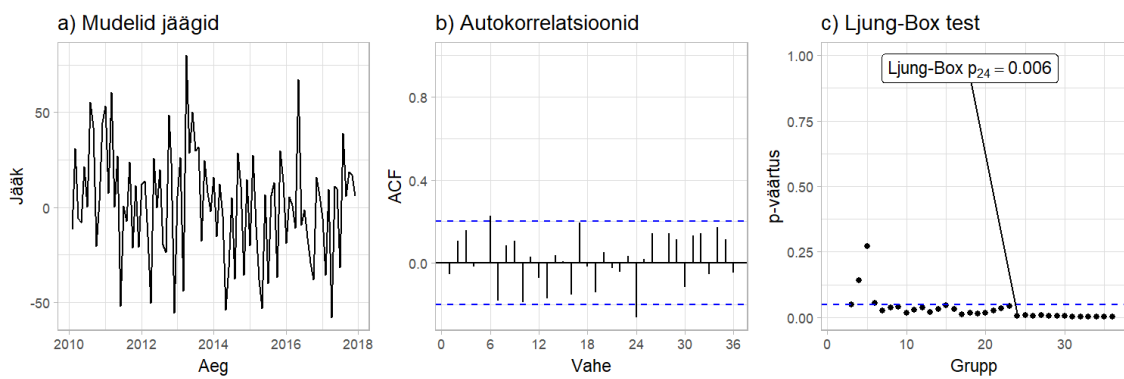
Joonis L2.2.2: Mudeli PED.m1.2 jääkide diagnostika (ARIMA(1,0,0)(2,0,0)₆)



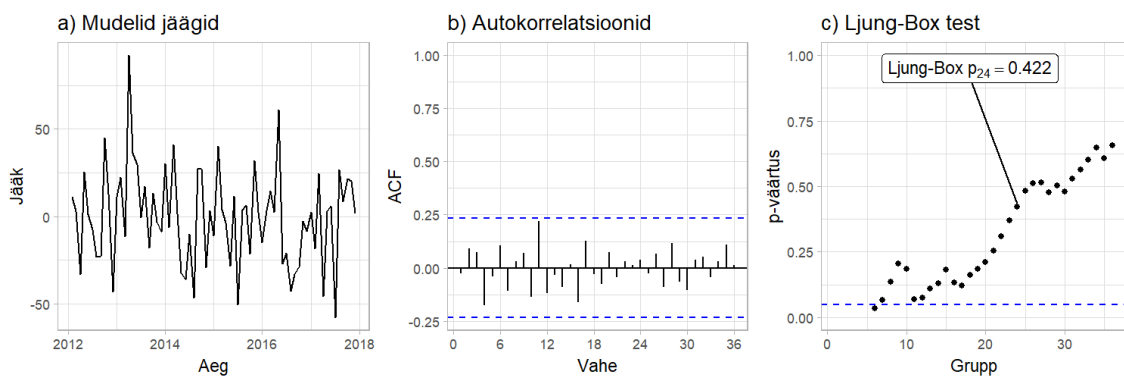
Joonis L2.2.3: Mudeli PED.m1.3 jääkide diagnostika (ARIMA(2,0,0)(1,0,0)₁₂)



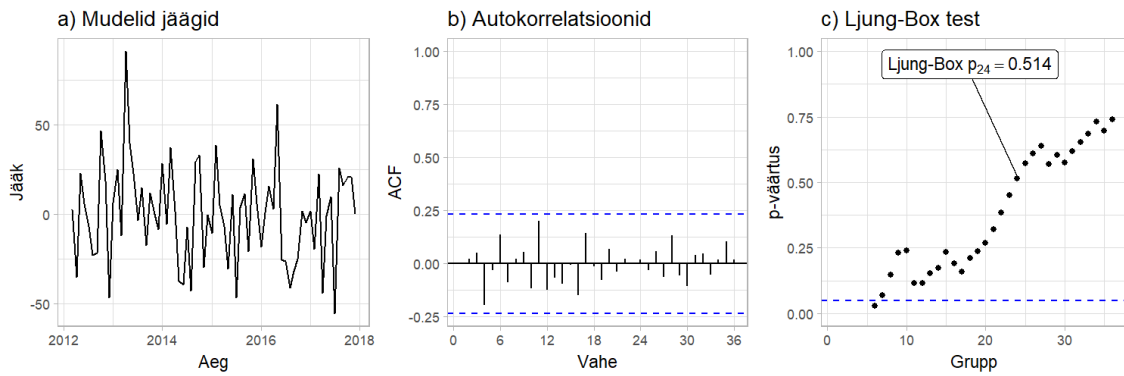
Joonis L2.2.4: Mudeli PED.m1.4 jääkide diagnostika ($\text{ARIMA}(2, 0, 0)(2, 0, 0)_6$)



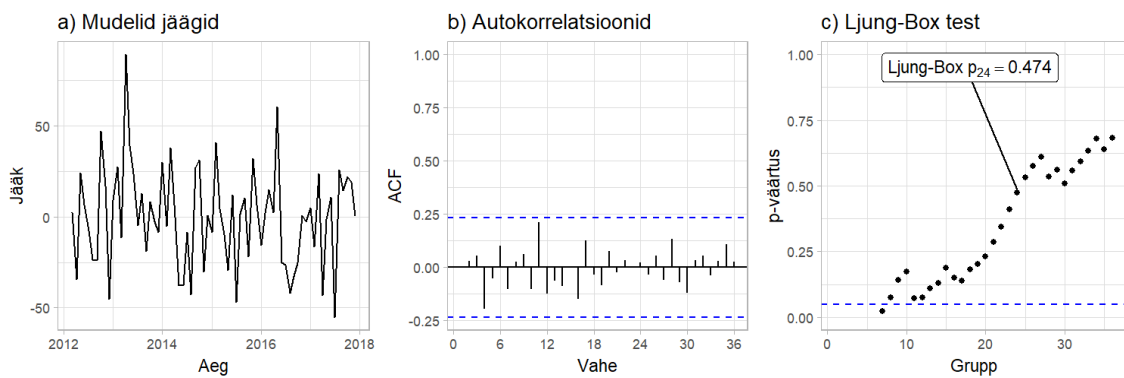
Joonis L2.2.5: Mudeli PED.m2.1 jääkide diagnostika ($\text{ARIMA}(1, 0, 0)(1, 1, 0)_{12}$)



Joonis L2.2.6: Mudeli PED.mx1.2 jääkide diagnostika (lineaarne regressioonimudel $\text{ARIMA}(1, 0, 0)(2, 0, 0)_6$ vigadega)

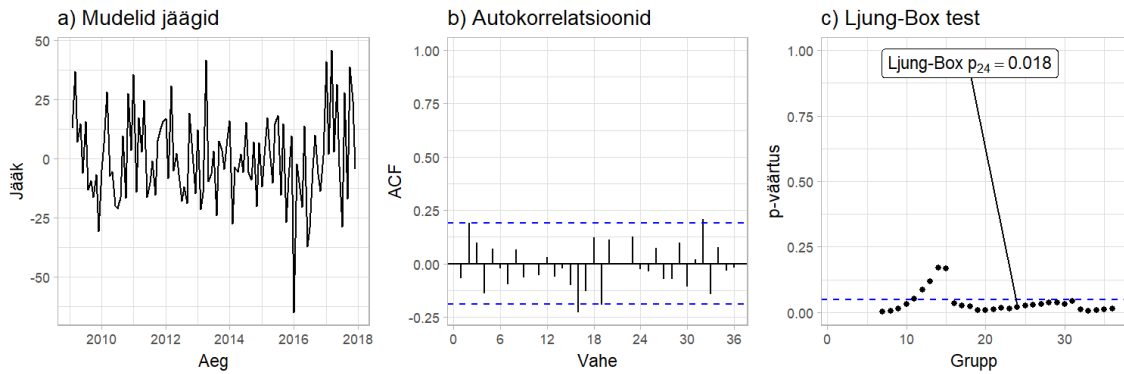


Joonis L2.2.7: Mudeli PED.mx1.3 jääkide diagnostika (lineaarne regressioonimudel $ARIMA(2, 0, 0)(1, 0, 0)_{12}$ vigadega)

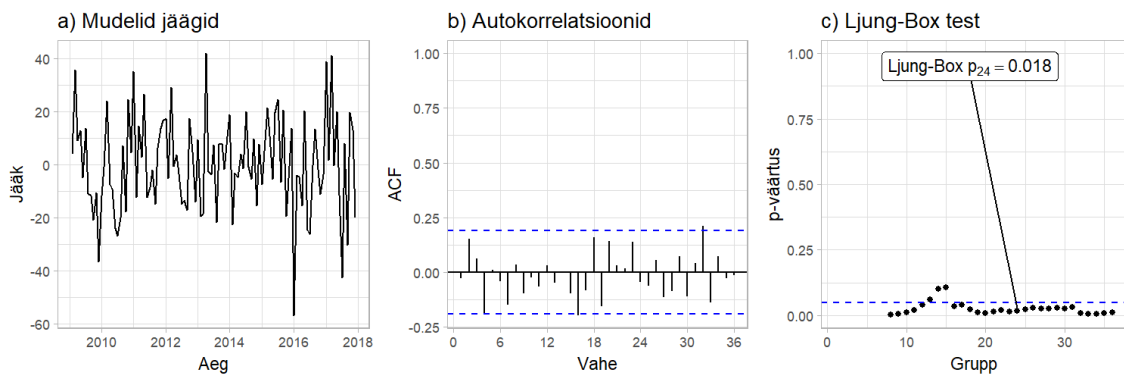


Joonis L2.2.8: Mudeli PED.mx1.4 jääkide diagnostika (lineaarne regressioonimudel $ARIMA(2, 0, 0)(2, 0, 0)_6$ vigadega)

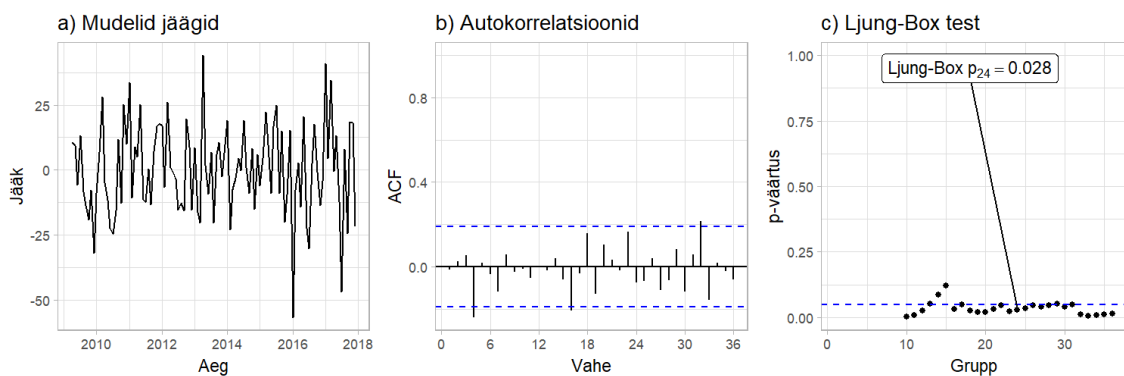
Lisa 2.3 Sisehaiguste täiendavaid mudelite jooniseid



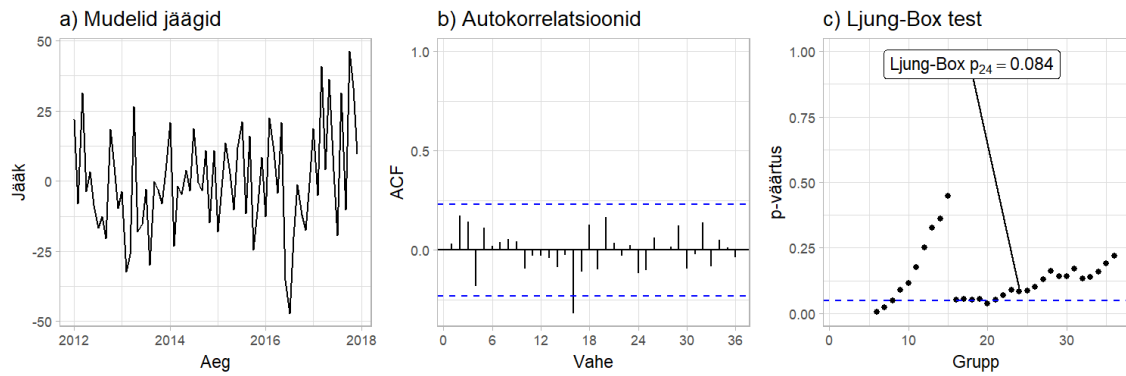
Joonis L2.3.1: Mudeli SISE.m1.1 jääkide diagnostika ($ARIMA(1, 0, 0)(2, 0, 0)_6$ trendi 1 eeldusel)



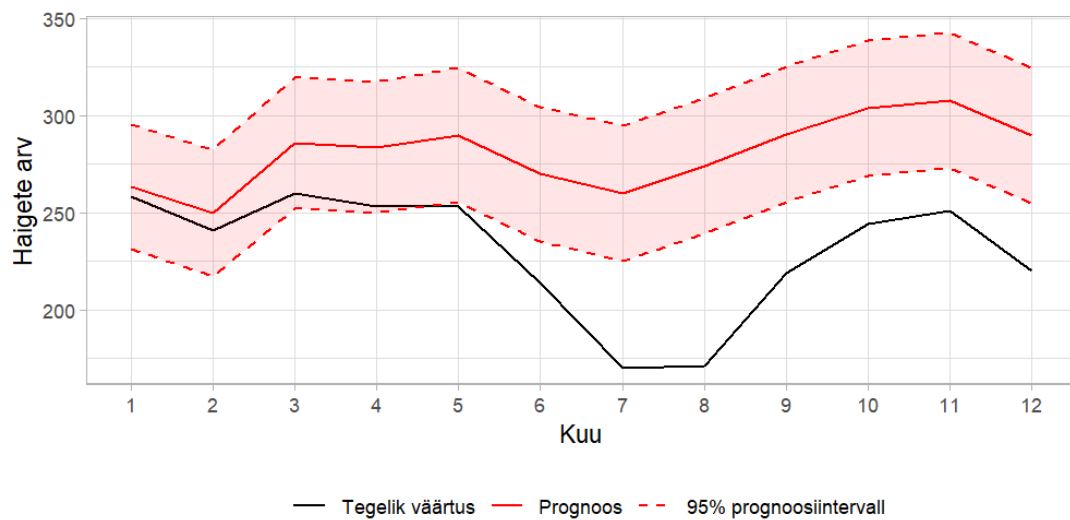
Joonis L2.3.2: Mudeli SISE.m2.1 jääkide diagnostika ($ARIMA(1, 0, 0)(2, 0, 0)_6$ trendi 2 eeldusel)



Joonis L2.3.3: Mudeli SISE.m2.2 jääkide diagnostika ($ARIMA(3, 0, 0)(2, 0, 0)_6$ trendi 2 eeldusel)

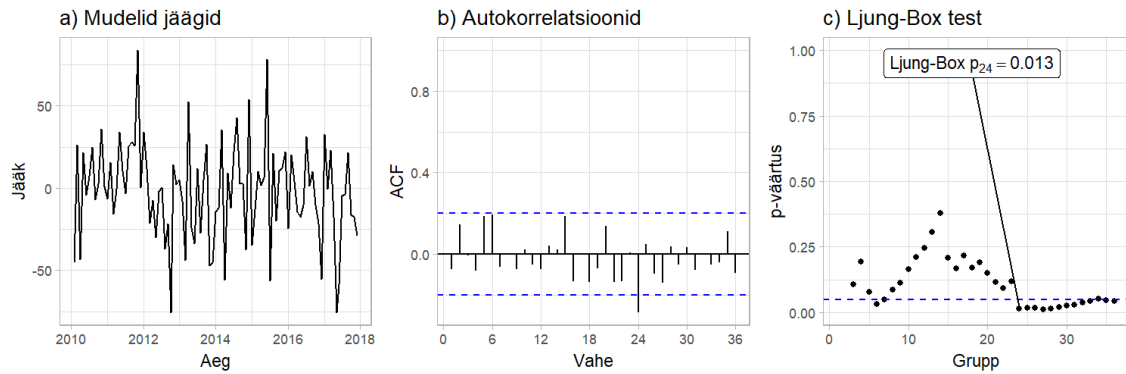


Joonis L2.3.4: Mudeli SISE.mx1.2 jääkide diagnostika (lineaarne regressioonimudel $ARIMA(0, 0, 1)(2, 0, 0)_6$ vigadega)

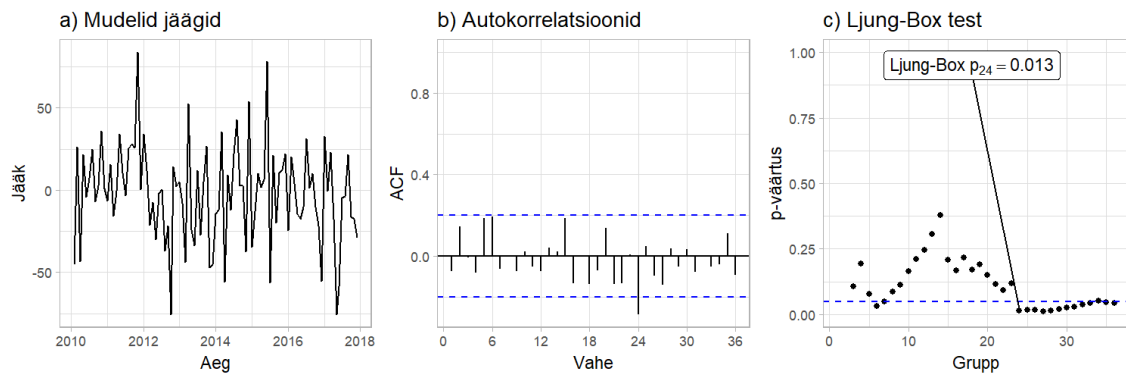


Joonis L2.3.5: Mudeli SISE.m2.3 prognoosid ($ARIMA(4, 0, 0)(2, 0, 0)_6$ trendi 2 eeldusel)

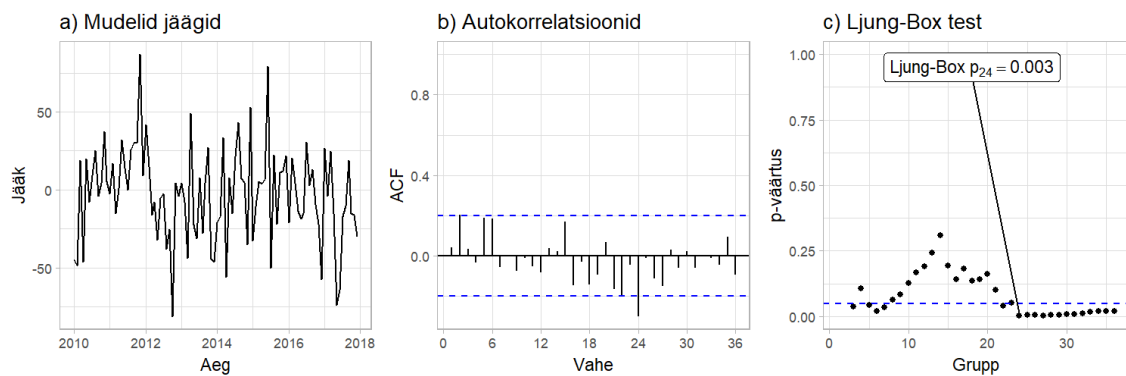
Lisa 2.4 Üld- ja plastikakirurgia täiendavaid mudelite jooniseid



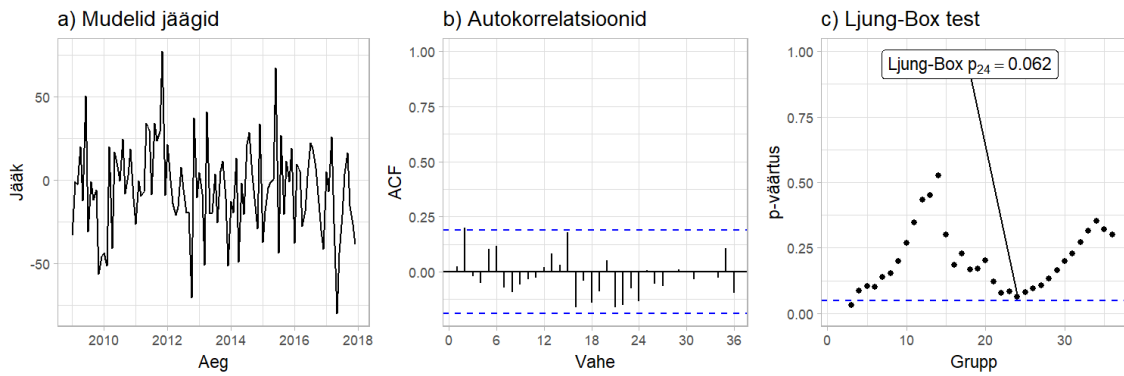
Joonis L2.4.1: Mudeli ÜPK.m1.5 jääkide diagnostika (ARIMA(2,0,0)(1,0,1)₁₂)



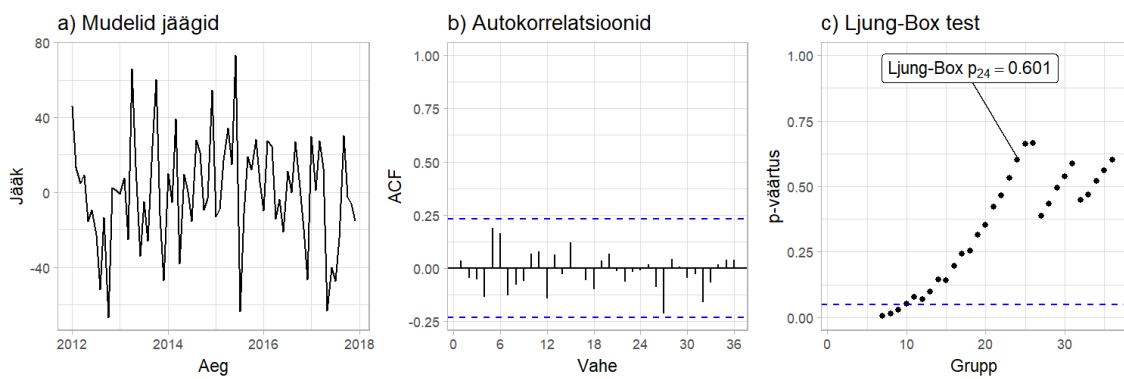
Joonis L2.4.2: Mudeli ÜPK.m2.1 jääkide diagnostika (ARIMA(1,0,0)(1,1,0)₁₂)



Joonis L2.4.3: Mudeli ÜPK.m2.2 jääkide diagnostika (ARIMA(0,0,1)(1,1,0)₁₂)

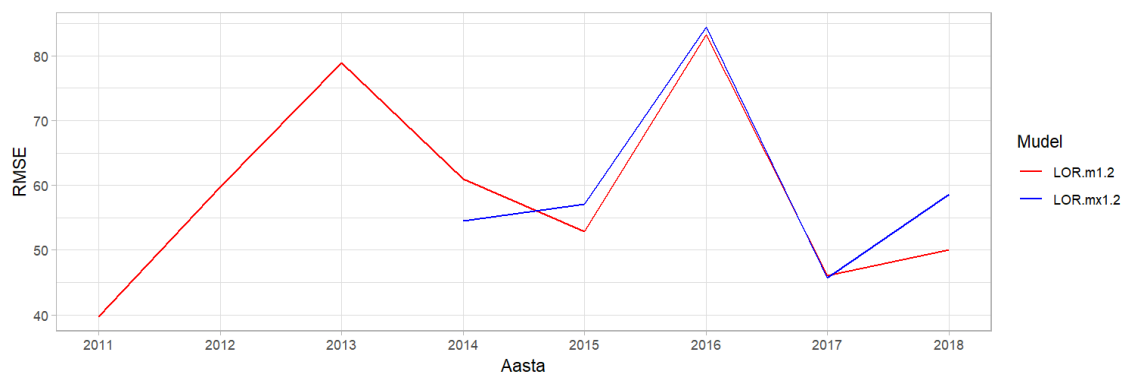


Joonis L2.4.4: Mudeli ÜPK.m2.4 jääkide diagnostika ($\text{ARIMA}(0, 0, 1)(0, 1, 1)_{12}$)

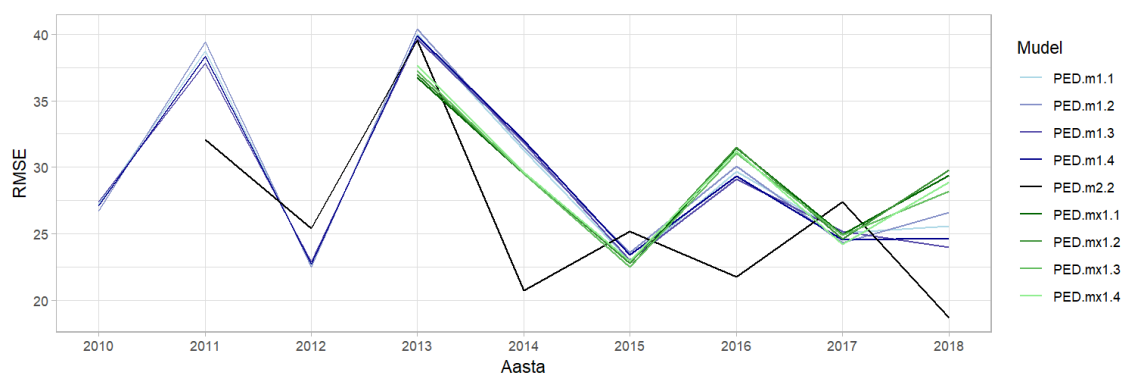


Joonis L2.4.5: Mudeli ÜPK.mx1.2 jääkide diagnostika (lineaarne regressioonimudel $\text{ARIMA}(0, 0, 3)(1, 0, 0)_{12}$ vigadega)

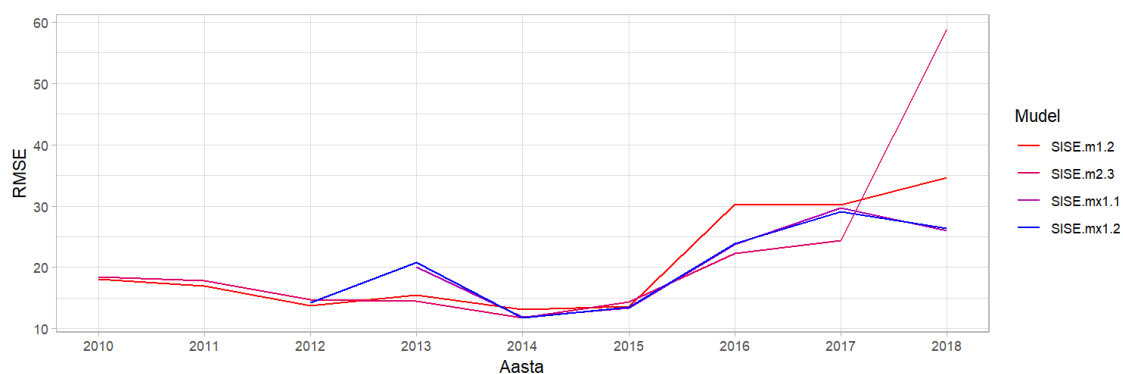
Lisa 3 Mudelite täpsusnäitajad



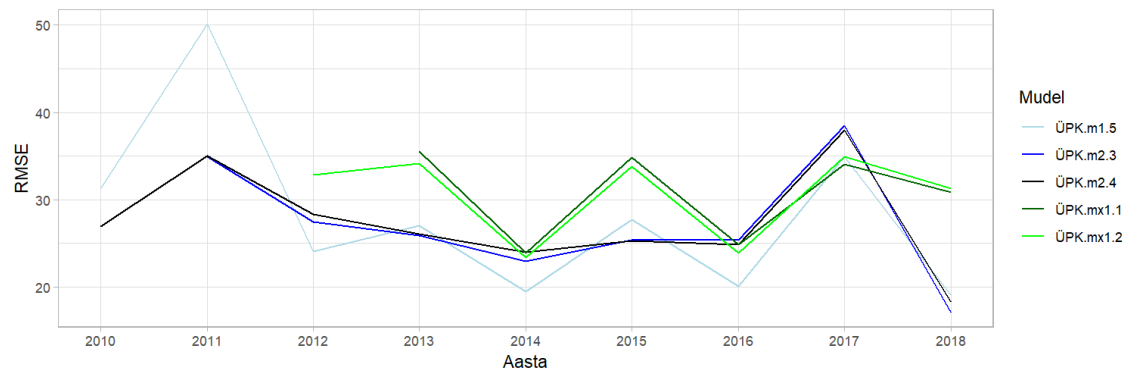
Joonis L3.1: Andmetega sobivate LORi mudelite 12-sammuliste prognooside ruutkeskmised vead treeningperioodis ning testaastal 2018



Joonis L3.2: Andmetega sobivate pediaatria mudelite 12-sammuliste prognooside ruutkeskmised vead treeningperioodis ning testaastal 2018



Joonis L3.3: Andmetega sobivate sisehaiguste mudelite 12-sammuliste prognooside ruutkeskmised vead treeningperioodis ning testaastal 2018



Joonis L3.4: Andmetega sobivate ÜPK mudelite 12-sammuliste prognooside ruutkeskmised vead treeningperioodis ning test aastal 2018

Lisa 4 Lõplike mudelite Ri väljundid

```
library(forecast)
```

```
# Otorinolarüngoloogia: parim ühemõõtmeline mudel
# Ajast sõltuv eksponentsiaalne trend
t = 1:120
m1 = nls(stat_lor_treening~c1+c2*exp(c3*t),
        start=c(c1=300,c2=100,c3=-0.1))
# Fikseeritud parameeter c3
c3 = coef(m1)[3]
# Lõplik mudel
LOR.m1.2 = Arima(stat_lor_treening, xreg=exp(c3*t), order=c(2,0,0),
                 seasonal=list(order=c(2,0,0), period=12),
                 method="CSS")
LOR.m1.2
```

```
## Series: stat_lor_treening
## Regression with ARIMA(2,0,0)(2,0,0)[12] errors
##
## Coefficients:
##          ar1          ar2          sar1          sar2  intercept  exp(c3 * t)
##          0.2550  -0.2377  0.3439  0.5323   235.0385    273.3358
## s.e.    0.0905   0.0898  0.0795  0.0839   93.3275    126.8429
##
## sigma^2 estimated as 2754:  part log likelihood=-657.1
```

```
# Otorinolarüngoloogia: parim mitmemõõtmeline mudel
# Lõplik mudel
LOR.mx1.2 = Arima(z_lor_treening, xreg=RRL_lor_treening, order=c(2,0,0),
                 seasonal=list(order=c(2,0,0), period=12),
                 method="CSS")
LOR.mx1.2
```

```
## Series: z_lor_treening
## Regression with ARIMA(2,0,0)(2,0,0)[12] errors
##
## Coefficients:
##          ar1          ar2          sar1          sar2  intercept  RRL_lor_treening
##          0.1906  -0.3097  0.2858  0.6361   -14.2427          -0.1618
## s.e.    0.1081   0.1073  0.0869  0.0920   399.3265           0.2657
##
## sigma^2 estimated as 2662:  part log likelihood=-462.88
```

```
# Pediaatria: parim ühemõõtmeline mudel
```

```
PED.m2.2 = Arima(stat_ped_treening, order=c(1,0,0),  
                 seasonal=list(order=c(0,1,1), period=12),  
                 method="CSS")
```

```
PED.m2.2
```

```
## Series: stat_ped_treening
```

```
## ARIMA(1,0,0)(0,1,1)[12]
```

```
##
```

```
## Coefficients:
```

```
##          ar1      sma1
```

```
##          0.2337 -0.6217
```

```
## s.e.    0.0945   0.0956
```

```
##
```

```
## sigma^2 estimated as 823.5:  part log likelihood=-515.27
```

```
# Pediaatria: parim mitmemõõtmeline mudel
```

```
PED.mx1.1 = Arima(z_ped_treening, xreg=RRL_ped_treening, order=c(1,0,0),  
                 seasonal=list(order=c(1,0,0), period=12),  
                 method="CSS")
```

```
PED.mx1.1
```

```
## Series: z_ped_treening
```

```
## Regression with ARIMA(1,0,0)(1,0,0)[12] errors
```

```
##
```

```
## Coefficients:
```

```
##          ar1      sar1  intercept  RRL_ped_treening
```

```
##          0.2858  0.6258   396.1071             -0.5099
```

```
## s.e.    0.1098  0.0824    72.4095             0.3864
```

```
##
```

```
## sigma^2 estimated as 663.1:  part log likelihood=-397.07
```

```
# Sisehaigused: parim ühemõõtmeline mudel
# Ajast sõltuva trend 1 eeldusel
t = 1:120
g = function(x) (x-97)*(x>=98)*(x<=102)+6*(x>102)
y_t = g(t)
SISE.ml.2 = Arima(stat_sise_treening, xreg=cbind(t, y_t), order=c(3,0,0),
                  seasonal=list(order=c(2,0,0), period=6),
                  method="CSS")
SISE.ml.2
```

```
## Series: stat_sise_treening
## Regression with ARIMA(3,0,0)(2,0,0)[6] errors
##
## Coefficients:
##          ar1      ar2      ar3      sar1      sar2  intercept      t      y_t
##          0.2507  0.2220  0.0262  0.2367  0.5119      135.569   1.3873 -10.5594
## s.e.      0.0906  0.0914  0.0987  0.0811  0.0862      62.931   0.6089   2.3717
##
## sigma^2 estimated as 302.2:  part log likelihood=-516.8
```

```
# Sisehaigused: parim mitmemõõtmeline mudel
SISE.mx1.1 = Arima(z_sise_treening, xreg=RRL_sise_treening,
                  order=c(1,0,0),
                  seasonal=list(order=c(2,0,0), period=6),
                  method="CSS")
SISE.mx1.1
```

```
## Series: z_sise_treening
## Regression with ARIMA(1,0,0)(2,0,0)[6] errors
##
## Coefficients:
##          ar1      sar1      sar2  intercept  RRL_sise_treening
##          0.2928  0.2841  0.4315  -243.6662           2.0616
## s.e.      0.1053  0.1024  0.1023    80.2715           0.3375
##
## sigma^2 estimated as 301.3:  part log likelihood=-363.42
```

```
# Üld- ja plastikakirurgia: parim ühemõõtmeline mudel
YPK.m2.3 = Arima(stat_yld_treening, order=c(1,0,0),
                 seasonal=list(order=c(0,1,1), period=12),
                 method="CSS")
YPK.m2.3
```

```
## Series: stat_yld_treening
## ARIMA(1,0,0)(0,1,1)[12]
##
## Coefficients:
##          ar1      sma1
##      0.3500 -0.7702
## s.e.  0.0893  0.0712
##
## sigma^2 estimated as 726.6:  part log likelihood=-508.51
```

```
# Üld- ja plastikakirurgia: parim mitmemõõtmeline mudel
YPK.mx1.1 = Arima(z_yld_treening, xreg=RRL_yld_treening, order=c(3,0,0),
                 seasonal=list(order=c(1,0,0), period=12),
                 method="CSS")
YPK.mx1.1
```

```
## Series: z_yld_treening
## Regression with ARIMA(3,0,0)(1,0,0)[12] errors
##
## Coefficients:
##          ar1      ar2      ar3      sar1  intercept  RRL_yld_treening
##      0.2244 -0.0452 -0.2518  0.5961   177.1679           0.1317
## s.e.  0.1071  0.1100  0.1074  0.0888   115.8581           0.5633
##
## sigma^2 estimated as 740.7:  part log likelihood=-401.86
```

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Hanna-Liisa Soll,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose "Haiglapatsientide arvu prognoosimine ARIMA tüüpi mudelitega", mille juhendajad on Krista Fischer, Raul Kangro ja Merje Tikk, reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.
2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 3.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.
3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Hanna-Liisa Soll

15.05.2019